

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds

(without alignments)
1251.566 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447

Sequence: 1 MGENDPPAVEAPFSEFSLFG.....TRVTSFUDWHEQMERD LKT 454

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

al number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2447	100.0	454	1	TMS3_HUMAN
2	865.5	35.4	490	1	TMS2_MOUSE
3	864.5	35.3	492	1	TMS2_HUMAN
4	704	28.8	455	1	TMS5_MOUSE
5	696	28.4	457	1	TMS5_HUMAN
6	664.5	27.2	417	1	HEPS_HUMAN
7	661.5	27.0	1069	1	ENTK_MOUSE
8	660.5	27.0	437	1	TMS4_HUMAN
9	660	27.0	1035	1	ENTK_BOVIN
10	657	26.8	1019	1	ENTK_HUMAN
11	654	26.7	1034	1	ENTK_PIG
12	650	26.6	416	1	HEPS_PIG
13	645	26.4	416	1	HEPS_RAT
14	570	23.3	638	1	KAL_RAT
15	561	22.9	638	1	KAL_MOUSE
16	540	22.1	638	1	KAL_HUMAN
17	533	21.7	855	1	ST14_HUMAN
18	532	21.6	324	1	PLMN_MOUSE
19	523.5	21.3	790	1	PLMN_PIG
20	520.5	21.3	761	1	NETR_MOUSE
21	520.5	21.3	338	1	PLMN_HORSE
22	520	21.3	1042	1	CORI_HUMAN
23	512.5	20.9	855	1	ST14_MOUSE
24	508.5	20.8	314	1	TEST_HUMAN
25	507.5	20.7	812	1	PLMN_MOUSE
26	507	20.7	437	1	ACRO_RAT
27	506	20.6	1113	1	CORI_MOUSE
28	504.5	20.6	810	1	PLMN_HUMAN
29	503.5	20.6	273	1	MCT7_MOUSE
30	503.5	20.6	436	1	ACRO_MOUSE
31	503	20.6	343	1	PLMN_SHEEP
32	501.5	20.5	810	1	PLMN_ERIEU
33	501.5	20.5	875	1	NETR_HUMAN

ALIGNMENTS

RESULT	ID	STANDARD	PRT	454 AA	ALIGNMENTS
1	TMS3_HUMAN				
1	TMS3_HUMAN				
AC	P57727				
DT	20-AUG-2001 (Rel. 40, Created)				
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE				
DE	TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).				
GN	TPRPS3 OR TADG12 OR ECHOS1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).				
RC	TISSUE=Ovarian carcinoma;				
TX	MEDLINE=20521358; PubMed=11068177;				
RA	Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,				
RA	Wang Y., Paroley T.H., O'Brien T.J.;				
RT	"Ovarian tumor cells express a novel multi-domain cell surface serine				
RT	protease.";				
RL	Biochim. Biophys. Acta 1502:337-350(2000).				
RL	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D), AND VARIANT ILE-53.				
RX	MEDLINE=20578749; PubMed=11137999;				
RA	Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrast R.,				
RA	Guppioni M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,				
RA	Younus F., Mohd S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,				
RA	Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,				
RA	Antonarakis S.E.;				
RT	"Insertion of beta-satellite repeats identifies a transmembrane				
RT	protease causing both congenital and childhood onset autosomal				
RT	recessive deafness.";				
RL	Nat. Genet. 27:59-63(2001).				
CC	-1- FUNCTION: PROBABLE PROTEASE.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	-1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND				
CC	TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED				
CC	ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.				
CC	-1- DISEASE: DEFECTS IN TPRPS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL				
CC	NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRIPSLIN FAMILY.				
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				

Query Match	Best Local Similarity	100.0%: Score 2447, DB 1; Length 454;
Matches 454; Conservative 100.0%; Mismatches 0; Indels 0; Gaps 0		
EMBL; AF201380; AAC37012.1; -		
DR EMBL; AB038158; BAB20077.1; -		
DR EMBL; AB038158; BAB20078.1; -		
DR EMBL; AB038158; BAB20079.1; -		
DR EMBL; AB038160; BAB20080.1; -		
DR MIM; 605511; -		
DR MIM; 601072; -		
DR MIM; 605316; -		
DR InterPro; IPR001314; Chymotrypsin.		
DR InterPro; IPR002172; LDL_recept_A.		
DR InterPro; IPR001190; SRCR.		
DR InterPro; IPR001354; Trypsin.		
DR Pfam; PF000057; ldl_recept_a; 1.		
DR Pfam; PF00530; SRCR; 1.		
DR Pfam; PF00089; trypsin; 1.		
DR SMART; SM00192; LDLa; 1.		
DR SMART; SM00202; SR; 1.		
DR SMART; SM00020; Tryp_Spc; 1.		
DR PROSITE; PS50068; LDLa_2; 1.		
DR PROSITE; PS50420; SRCR_1; FALSE_NEG.		
DR PROSITE; PS50287; SRCR_2; 1.		
DR PROSITE; PS50240; TRYPsin_DOM; 1.		
DR PROSITE; PS00134; TRYPsin_HIS; 1.		
DR PROSITE; PS00135; TRYPsin_SER; 1.		
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Deafness;		
KW Alternative splicing; Polymorphism.		
KM CYTOPLASMIC (POTENTIAL).		
KT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)		
FT TRANSMEM 49 69		
FT DOMAIN 1 48		
FT DOMAIN 70 454		
FT DOMAIN 72 108		
FT DOMAIN 109 205		
FT DOMAIN 217 454		
FT ACT_SITE 257 257		
FT ACT_SITE 304 304		
FT ACT_SITE 401 401		
FT SITE 216 217		
FT DISULFID 73 85		
FT DISULFID 79 98		
FT DISULFID 92 107		
FT DISULFID 129 194		
FT DISULFID 142 204		
FT DISULFID 207 324		
FT DISULFID 242 258		
FT DISULFID 338 407		
FT DISULFID 370 386		
FT DISULFID 397 425		
FT CARBOHD 221 221		
FT VARSPLIC 1 127		
FT VARSPLIC 318 454		
FT VARSPLIC 261 293		
FT VARSPLIC 294 454		
FT VARIANT 53 53		
FT CONFLICT 46 54		
FT CONFLICT 90 90		
FT CONFLICT 350 350		
FT CONFLICT 369 395		
FT CONFLICT 427 427		
FT SEQUENCE 454 AA; 49404 MW; 57BC3678F7D6AF CRC64;		

OY	1	MGENDPVAVEAPFEFERSLFGIDLDKITSVPAPDADAVALAAILSLPLKEPPIIVIGIILI	60
Db	1	MGENDPVAVEAPFEFERSLFGIDLDKITSVPAPDADAVALAAILSLPLKEPPIIVIGIILI	60
OY	61	LALATGLIGHEDCSGKYRCRSSFECIELLARCDGVSDCKDGEDEYRCVAVGGONAVLOVF	120
Db	61	LALATGLIGHEDCSGKYRCRSSFECIELLARCDGVSDCKDGEDEYRCVAVGGONAVLOVF	120
OY	121	TAAASKTKCSDMKGHYANNAACAOLGPPSYSSDNLRVSLEGOFREEEYSIDLHLPDK	180
Db	121	TAAASKTKCSDMKGHYANNAACAOLGPPSYSSDNLRVSLEGOFREEEYSIDLHLPDK	180
OY	181	VTAALHSYTYVEGCASGVHYVTLOCTACGHRGRGSRSRTIVSGNNLSLSOMPMOASLOFOGYH	240
Db	181	VTAALHSYTYVEGCASGVHYVTLOCTACGHRGRGSRSRTIVSGNNLSLSOMPMOASLOFOGYH	240
OY	241	LCGGSVITPLMTITTAACHCYVDLYLPKSWTIOVLQVLSILDNPAPSHLYEKITYVSKYKKRR	300
Db	241	LCGGSVITPLMTITTAACHCYVDLYLPKSWTIOVLQVLSILDNPAPSHLYEKITYVSKYKKRR	300
OY	301	LGNDIALMLKLAGPLTFNEMIQPVCLRPSENFEPDGKYCVTSCMGARIEDGAGSAYLVLHA	360
Db	301	LGNDIALMLKLAGPLTFNEMIQPVCLRPSENFEPDGKYCVTSCMGARIEDGAGSAYLVLHA	360
OY	361	AVPLLNSNKICNHRYVGIIISPSMLCAGYTGTGVDSCOGSGSPPLYVCORRLMKLVGATS	420
Db	361	AVPLLNSNKICNHRYVGIIISPSMLCAGYTGTGVDSCOGSGSPPLYVCORRLMKLVGATS	420
OY	421	FGIGCAEVNKPQVYTRVTSFLDMITHQMERDLKT	454
Db	421	FGIGCAEVNKPQVYTRVTSFLDMITHQMERDLKT	454
<hr/>			
RESULT 2			
ID	TMS2_MOUSE	STANDARD;	PRT; 490 AA.
AC	09JUI08; 09JKC4; 09OY82;		
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	TRANSMEMBRANE PROTEASE, SERINE 2 (BC 3.4.21.-) (EPITHELLASIN) (PLASMIC		
GN	TPRSS2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCB1_TextID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.		
RC	STRAIN-BALB/C;		
RX	MEDLINE=21104370; PubMed=11169526;		
RA	Vaalala M.H., Porvari K.S., Kellonkumpu S., Kyllonen A.P., Vihko P.T.;		
RT	"Expression of transmembrane serine protease TMPRSS2 in mouse and		
RT	human tissues.";		
RL	J. Pathol. 193:134-140(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Han J., Kim S.;		
RT	"Putative transmembrane protease X.";		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C;		
RA	Jacquinet E.J., Rao N.V., Rao G.N., Hoidl J.R.;		
RT	"A novel mosaic serine protease, epitheliasin.";		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.		
CC	- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.		
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRIPSTIN FAMILY.		
CC	- SIMILARITY: CONTAINS 1 SRCR DOMAIN.		
CC	- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.		

PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PA (GENETECH) GENENTECH INC.
 XX
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
 PI Ferrara N, Flier A, Fong S, Gao W, Gerber H, Gertszen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 DR WPI, 2000-611443/58.
 DR N-PSDB; C78475.
 PI Novel PRO polypeptides and polynucleotides used in detection methods,
 PI to target bioactive molecules to specific cells, and to modulate
 PI cellular activities -
 PS Claim 12; Fig 28; 636pp; English.
 XX
 XX C78458 to C78599 represent polynucleotide and EST (expressed sequence
 CC tag) sequences which encode secreted or transmembrane PRO polypeptides.
 CC The PRO polynucleotides and polypeptides have cytosolic activity.
 CC The polynucleotides and polypeptides can be used for detecting the
 CC presence of PRO polypeptides in samples, for linking bioactive molecules
 CC to cells and for modulating biological activities of cells, using the
 CC polypeptides for specific targeting. The polypeptide targeting can be
 CC used to kill the target cells, e.g. for the treatment of cancers. The
 CC polypeptide pairs provide specific targeting of bioactive molecules to
 CC cells. C78600 to C78987 represent PCR primers and probes used in the
 CC isolation of the PRO polynucleotide sequences.
 CC
 XX Sequence 453 AA;
 XX
 Query Match 99.9%; Score 2441; DB 21; Length 453;
 Best Local Similarity 99.8%; Pred. No. 2.6e-188;
 Matches 452; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
 ID Y41694
 Y41694 standard; Protein; 452 AA.
 AC Y41694;
 XX
 DT 07-DEC-1999 (first entry)
 XX
 DE Human PRO382 protein sequence.
 DE
 XX Human: PRO: EST: expressed sequence tag; PCR primer; hybridisation;
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 KW secreted protein; transmembrane protein.
 OS Homo sapiens.
 XX
 PN WO9446281-A2.
 XX
 PD 16-SEP-1999.
 XX
 PF 08-MAR-1999;
 XX
 PR 10-MAR-1998; 98US-0077450.
 PR 11-MAR-1998; 98US-0077632.
 PR 11-MAR-1998; 98US-0077641.
 PR 11-MAR-1998; 98US-0077649.
 PR 12-MAR-1998; 98US-0077791.
 PR 13-MAR-1998; 98US-0078004.
 PR 17-MAR-1998; 98US-0040220.
 PR 20-MAR-1998; 98US-0078886.
 PR 20-MAR-1998; 98US-0078910.
 PR 20-MAR-1998; 98US-0078936.
 PR 20-MAR-1998; 98US-0078939.
 PR 25-MAR-1998; 98US-0079294.
 PR 25-MAR-1998; 98US-0079656.
 PR 26-MAR-1998; 98US-0079663.
 PR 27-MAR-1998; 98US-0079664.
 PR 27-MAR-1998; 98US-0079689.
 PR 27-MAR-1998; 98US-0079728.
 PR 27-MAR-1998; 98US-0079786.
 PR 30-MAR-1998; 98US-0079920.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080107.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.
 PR 01-APR-1998; 98US-0080327.
 PR 01-APR-1998; 98US-0080328.
 PR 01-APR-1998; 98US-0080333.
 PR 01-APR-1998; 98US-0080334.
 PR 01-APR-1998; 98US-0081049.
 PR 08-APR-1998; 98US-0081070.
 PR 08-APR-1998; 98US-0081071.
 PR 09-APR-1998; 98US-0081195.
 PR 09-APR-1998; 98US-0081203.
 PR 09-APR-1998; 98US-0081229.
 PR 09-APR-1998; 98US-0081817.
 PR 15-APR-1998; 98US-0081838.
 PR 15-APR-1998; 98US-0081952.
 PR 15-APR-1998; 98US-0081955.
 PR 21-APR-1998; 98US-0082568.
 PR 21-APR-1998; 98US-0082569.
 PR 22-APR-1998; 98US-0082700.
 PR 22-APR-1998; 98US-0082704.
 PR 22-APR-1998; 98US-0082804.
 PR 23-APR-1998; 98US-0082767.
 PR 23-APR-1998; 98US-0082796.
 PR 27-APR-1998; 98US-0083336.
 PR 28-APR-1998; 98US-0083322.
 PR 29-APR-1998; 98US-0083392.
 PR 29-APR-1998; 98US-0083495.
 PR 29-APR-1998; 98US-0083952.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 15.91 seconds
(without alignments)
642.143 Million cell updates/sec

Title: US-09-846-512-12
2447
Sequence: 1 MGENDPPEAVEAFSFRSLFG.....TRVTSFLDWIHEQMERDLKT 454

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2375	97.1	454	US-09-518-046-2	Sequence 2, Appl1
2	2195.5	89.7	455	US-09-261-416-2	Sequence 2, Appl1
3	1354	55.3	294	US-09-518-046-4	Sequence 4, Appl1
4	875.5	35.8	492	US-09-342-749-2	Sequence 2, Appl1
5	851.5	34.8	157	US-09-518-046-23	Sequence 23, Appl1
6	686.5	28.1	283	US-08-807-151-1	Sequence 1, Appl1
7	660.5	27.0	435	US-09-008-271A-6	Sequence 6, Appl1
8	660	27.0	798	US-08-200-900A-2	Sequence 2, Appl1
9	660	27.0	798	US-09-000-846-2	Sequence 2, Appl1
10	655.5	26.8	416	US-09-027-337-3	Sequence 3, Appl1
11	576	23.5	255	US-08-944-483-67	Sequence 67, Appl1
12	571	23.3	255	US-08-681-151-3	Sequence 3, Appl1
13	570	23.3	638	US-08-508-448C-25	Sequence 25, Appl1
14	564.5	23.1	418	US-08-508-448C-19	Sequence 19, Appl1
15	533	21.8	855	US-09-027-337-2	Sequence 2, Appl1
16	533	21.6	248	US-08-944-483-63	Sequence 63, Appl1
17	529.5	21.5	235	US-08-807-151-3	Sequence 3, Appl1
18	523	21.4	98	US-09-518-046-17	Sequence 17, Appl1
19	518	21.2	235	US-08-944-483-65	Sequence 65, Appl1
20	509.5	20.8	546	5200340-6	Patent No. 5200340
21	508.5	20.8	314	US-09-008-271A-3	Sequence 3, Appl1
22	507.5	20.7	812	US-08-248-629A-1	Sequence 1, Appl1
23	507.5	20.7	812	US-08-451-932-1	Sequence 1, Appl1
24	507.5	20.7	812	US-08-452-260-1	Sequence 1, Appl1
25	507.5	20.7	812	US-08-326-785-1	Sequence 1, Appl1
26	507.5	20.7	812	US-08-612-788-1	Sequence 1, Appl1
27	507.5	20.7	812	US-08-612-788-1	Sequence 1, Appl1

28	507.5	20.7	812	US-08-605-598B-1	Sequence 1, Appl1
29	507.5	20.7	812	US-08-429-743-1	Sequence 1, Appl1
30	507.5	20.7	812	US-08-866-743-1	Sequence 1, Appl1
31	507.5	20.7	812	US-09-066-028-1	Sequence 1, Appl1
32	507.5	20.7	812	PCT-US93-05107-1	Sequence 1, Appl1
33	507	20.7	814	US-08-750-711-1	Sequence 1, Appl1
34	505.5	20.7	791	US-08-643-219-1	Sequence 1, Appl1
35	505.5	20.7	791	US-08-851-350-1	Sequence 1, Appl1
36	504.5	20.6	790	US-08-469-486-54	Sequence 54, Appl1
37	504.5	20.6	790	US-08-469-658-54	Sequence 54, Appl1
38	504.5	20.6	791	US-09-131-995-1	Sequence 1, Appl1
39	504.5	20.6	791	US-08-832-087B-1	Sequence 1, Appl1
40	504.5	20.6	791	US-09-132-154-1	Sequence 1, Appl1
41	504.5	20.6	810	US-07-854-603-2	Sequence 2, Appl1
42	504.5	20.6	810	US-08-147-000B-29	Sequence 29, Appl1
43	504.5	20.6	810	US-09-086-514-1	Sequence 1, Appl1
44	504.5	20.6	810	5200340-8	Patent No. 5200340
45	503.5	20.6	273	US-08-978-404B-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: In Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADS-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match	97.1%	Score 2375;	DB 4;	Length 454;
Best Local Similarity	97.6%	Pred No. 1.5e-224;		
Matches 444;	Conservative	2;	Mismatches	7;
			Indels	2;
			Gaps	2;
1	MGENDPPEAVEAFSFRSLFGDLDLKISPAVDADAVAAQILSLPLKPF-PIVIGIIL	59		
1	MGENDPPEAVEAFSFRSLFGDLDLKISPAVDADAVAAQILSLPLKPF-PIVIGIIL	59		
60	ILALAIIGLHHDGSKRYKCRSSFKIELIARCDNRSDCKDDEDEYRCYRGGVANAAY	119		
61	ILALAIIGLHHDGSKRYKCRSSFKIELIARCDNRSDCKDDEDEYRCYRGGVANAAY	119		
120	PTAASWKTCSDMKGHHYANVACAOLGFPYSIDLRVSSLEGOFRFEVSIIDHLLPPD	179		
121	PTAASWKTCSDMKGHHYANVACAOLGFPYSIDLRVSSLEGOFRFEVSIIDHLLPPD	180		
180	KYTAALHSHYVREGCASGHVVTLOCTACGHRGYSRSRIYGNMSLLSQWPAQSLDPOGY	239		
181	KYTAALHSHYVREGCASGHVVTLOCTACGHRGYSRSRIYGNMSLLSQWPAQSLDPOGY	240		
240	HICGGSVITPLILITTAHGVYDLYLPKSWTIOGVGLVSLDNRAPSHLYEKTYYHSKYPK	299		
241	HICGGSVITPLILITTAHGVYDLYLPKSWTIOGVGLVSLDNRAPSHLYEKTYYHSKYPK	300		
300	RLGNDIALMKLAGPLTFNEMIOVCLPNSEENFPDGKVCWTGSGMGATEDGADAPVLNH	359		

Db 301 RLGNDAIIMKLAGPLTFNEMIOVCLPNSSENEPDGKVCWTSGWATEDS-CDASPYLNH 359
QY 360 AAVPLISNKCINHRDYGGLIISPSMCAAGLTGCVSDSCDDSGPLVCORRLMKLVAT 419
Db 360 AAVPLISNKCINHRDYGGLIISPSMCAAGLTGCVSDSCDDSGPLVCORRLMKLVAT 419
QY 420 SFGIGCAEVNKPVGYYTFTVTSFLDMIHQEMERDLKT 454
Db 420 SFGIGCAEVNKPVGYYTFTVTSFLDMIHQEMERDLKT 454

RESULT 2
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
; Patent No. 6291663
US-09-261-416-2

Query Match 89.7%; Score 2195.5; DB 4; Length 455;
Best Local Similarity 91.5%; Pred. No. 6e-207;
Matches 421; Conservative 5; Mismatches 23; Indels 11; Gaps 4;

QY 1 MGNDPPAVAPSPFSLFLDLKISPAVDADVAQAOLISLLPKFF-PIIVIGIAL 59
Db 1 MGNDPPAVAPSPFSLFLDLKISPAVDADVAQAOLISLLPKFFVFSQSSSLGIAL 60
QY 60 ILALAIGLIGHFDCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 119
Db 61 ILALAIGLIGHFDCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 120
QY 120 FTAASWKTMCSDMKGHYANVACAOLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 179
Db 121 FTAASWKTMCSDMKGHYANVACAOLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 180
QY 180 KYTALHHSYVVRGSCAGHYVTLQCTACGHRGYSRIVGNNLSLSQWQASLOFOGY 239
Db 181 KYTALHHSYVVRGSCAGHYVTLQCTACGHRGYSRIVGNNLSLSQWQASLOFOGY 240
QY 240 HLCGGSVITPLMTITTAHCYVDYLPRKSWTIQYGLVSLDNPAPSHLYEKIVYHKKRPK 299
Db 241 HLCGGSVITPLMTITTAHCYVDYLPRKSWTIQYGLVSLDNPAPSHLYEKIVYHKKRPK 300
QY 300 RLGNDAIIMKLAGPLTFNEMIOVCLPNSSENEPDGKVCWTSGWATEDSAGDASPYLNH 359
Db 301 RLGNDAIIMKLAGPLTFNEMIOVCLPNSSENEPDGKVCWTSGWATEDS-CDASPYLNH 359
QY 360 AAVPLISNKCINHRDYGGLIISPSMCAAGLTGCVSDSCDDSGPLVCORRLMKLVAT 419
Db 360 AAVPLISNKCINHRDYGGLIISPSMCAAGLTGCVSDSCDDSGPLVCORRLMKLVAT 419
QY 415 LVGATSEFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 454
Db 416 LVGATSEFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 455

RESULT 3
US-09-518-046-4
; Sequence 4, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 4
; LENGTH: 294
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
US-09-518-046-4

Query Match 55.3%; Score 1354; DB 4; Length 294;
Best Local Similarity 88.1%; Pred. No. 9.7e-125;
Matches 258; Conservative 7; Mismatches 20; Indels 8; Gaps 2;

QY 1 MGNDPPAVAPSPFSLFLDLKISPAVDADVAQAOLISLLPKFF-PIIVIGIAL 59
Db 1 MGNDPPAVAPSPFSLFLDLKISPAVDADVAQAOLISLLPKFFVFSQSSSLGIAL 60
QY 60 ILALAIGLIGHFDCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 119
Db 61 ILALAIGLIGHFDCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 120
QY 120 FTAASWKTMCSDMKGHYANVACAOLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 179
Db 121 FTAASWKTMCSDMKGHYANVACAOLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 180
QY 180 KYTALHHSYVVRGSCAGHYVTLQCTACGHRGYSRIVGNNLSLSQWQASLOFOGY 239
Db 181 KYTALHHSYVVRGSCAGHYVTLQCTACGHRGYSRIVGNNLSLSQWQASLOFOGY 240
QY 240 HLCGGSVITPLMTITTAHCYVDYLPRKSWTIQYGLVSLDNPAPSH 285
Db 241 HLCGGSVITPLMTITTAHCYVDYLPRKSWTIQYGLVSLDNPAPSH 293

RESULT 4
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-342-749-2


```

Db      250 TSMGSCAKAYRPGYGNWVFETDWTIRQMRAD 282

RESULT          7
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
            Hillman, Jennifer L.
            Yue, Henry
            Guegler, Karl J.
            Corley, Neil C.
            Tang, Tom Y.
            Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 137018
SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match                27.0%, Score 660.5; DB 4; Length 435;
Best Local Similarity     35.4%; Pred. No. 1.6e-56;
Matches 157; Conservative 73; Mismatches 150; Indels 63; Gaps 17;

    44 LPKKEPPIIVIGIILITLALA-----ICLGTHPCDSGRYRRSSFKC---IELLAR--- 91
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      22 IPMEFFRVVGPDIITALLSLASIIIVVLIKVILP--KY----YFLCGQPFIHPRKOL 74
    92 CDGVSDCKDEDEYRC-----VRVGONAVLVGFYTA--SMKTMGSDDMKGHYA 138
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      75 CDGEIDCPGLGEDBEHCYVSFPBGPAVAAYRLSKDRSTLYVDATGNGWFACDNFEALAA 134
    139 NVACAQLGFPSPYVSSDNLNRVSSLLEGQFREEFYSIDLHLLPDDKVYVALAHHSVVYREG--CA 195
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      135 ETACROMGYSS-----KPTFRAVEICGPDODLDVVEITEENSQELRRRNSSG PCT 182
    196 SGHVYTLOCTACGHRRGYSRSRIYVGGMNLSLQMPNQASIQFGYHLCCGSVTTPLMITIA 255
       ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

```

Db      183 SCSLSTSLHCLAGGEEL- KTRPRVVGGEASVDSPMQVSIOYDKQHVCGSSILDPHMVLTA 241
Oy      256 AHCV---IDLYLPKRSWTLQVGLVSLDNPAPSHLVKYIV---YHSKYPKRLGNIDIALMK 309
          |||             : : : : : |||             : : |||            |
Db      242 AHCFKRHTDVF--NMKVVRAGSKDL--GSPSLAVAKIIIEFNPMY-PK--DNIDIALMK 293
          |||             : : : : : |||             : : |||            |
Oy      310 LAGPLTFENMIQPVCLPMPSENEFPDGKVCWTSGMGATEDGADASPVLNHAVALPSLNKI 369
          | |||| : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Db      294 LQPLTFESGVTPICLLPFDELTATPLMTIGMCFTRONGSKMSDILLQASVOYIDSR 353
          |||             : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Oy      370 CNHRDYGGIISPSMLCAGYLTVGDSCQGGSDSGPLVCQERRLMKLVGATSFGICAEYN 429
          |||             : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Db      354 CANADAYAGEVTEKKMCAGIEPGVDTCQGSGGGLMYQSQQ-WHHVGVISWGCGGPS 412
          |||             : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Oy      430 KPGVTRTVTFELDMIHEDMERDL 452
          ||||| : : : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Db      413 TPGYTTKVSAYLNLWYNWKREL 435
          ||||| : : : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |

RESULT      8
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Melnert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-200-900A-2

Query Match      27.0%; Score 60; DB 1; Length 798;
Best Local Similarity 36.9%; Pred. No. 4.3e-56;
Matches 146; Conservative 70; Mismatches 146; Indels 34; Gaps 15;

Oy      66 GLGINFDC-SGRYYRORSRKCIETLIARDDGVSDCKDGDENVKCYRV---GCQMNVLOYVF 120
          ||||| : : : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Db      415 GLGIPEPCKEPNFKCKDG-ECIPLVNLCDDGPBHCXKDSDEAHCVLFLNGTDDSSGLVQR 473
          ||||| : : : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Oy      121 TAAASKTKSCDDMMKHVYNAVACAOIGFPSYVSSDWLRVYSLEGQRFEEFYSIDHLPPDK 180
          |||             : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Db      474 IQSINHVAACAEMWITQIIDDYCOLLGLET--GNSSVPFFSTIG---GPYVNLN----- 521
          |||             : : : ||| : | |||| : : : ||| : | |||| : : : ||| : |
Oy      181 VTALHNHYVV--RECCASGHVVTLOCT--ACGHR--RGYSSRIYCGNNSSLSONPQNAS 233

```

DB 522 -TAPNSLLTFSQCLEDSLLLCNYSKSGCKLVTQVSKRYTSGSRSRGAMWVYA 580
OY 234 LOPQGYHLCGGSVITPLWITTAHCVYDLYL--PKSWTIOVGL--VSLDNP-APSHLYEK 289
DB 581 LYFDDQVCGASLVSRDMLVSAHCVYGRNMPSPKKAVALGLHMASNLTSPQIETRLIDQ 640
OY 290 IYHSHYKRRKRLGNIDIALMKLAGPLTFNEMIOVCLPNSSENFPPQKCVWTSWGATTELG 349
DB 641 IYINPHYKRRKRNNDIAMHMLEKMYNYDYIOPICLPENOVFPGRICISIAWGAL-1Y 699
OY 350 AGDASPVLNHAAPLISNKCINHR--DVGIIISPSMLCAGYLTGVDSCQDGSGLVYCO 408
DB 700 QGSTADVIGEADVPILSNKCCQOQMPHYN--ITENMVCAGYEAGVDSQDGSGLVYCO 757
OY 409 ERRLKLVGATSFQIGCAEVNKPQYTRVTSFLDWI 444
DB 758 ENRMVLGAVTSFGYOCALPNRPYARVPRTFEMI 793

RESULT 9

PCT-US94-00616-2
Sequence 2, Application PC/TUS9400616
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 33
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00616
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 27.0%, Score 660; DB 5; Length 798;
Best Local Similarity 36.9%; Pred. No. 4.3e-56;

Matches 146; Conservative 70; Mismatches 146; Indels 34; Gaps 15;

OY 66 GLGIFDC-SGKYRCRSSFKCIELARCDGVSDCKGDEHYCVRY---GGQNAVLYOF 120
DB 415 GLGIFDC-SGKYRCRSSFKCIELARCDGVSDCKGDEHYCVRY---GGQNAVLYOF 120
OY 121 TAASRKTWCSDDMKGHYANVACAQGFPSYSSDNLRVSSLEGQRFEEFVSIDHLLPDK 180
DB 474 IQSIHVAACAEWMTQIDDDVQQLGLGT--GNSSVPFFSTG---GRYVNLN----- 521
OY 181 VTALHSHYV--REGSCAGHVVTLQCT--ACGHR--RGYSSRIVGNNMSLISQMPWQAS 233
DB 522 -TAPNSLLTFSQCLEDSLLLCNYSKSGCKLVTQVSKRYTSGSRSRGAMWVYA 580
OY 234 LOPQGYHLCGGSVITPLWITTAHCVYDLYL--PKSWTIOVGL--VSLDNP-APSHLYEK 289
DB 581 LYFDDQVCGASLVSRDMLVSAHCVYGRNMPSPKKAVALGLHMASNLTSPQIETRLIDQ 640
OY 290 IYHSHYKRRKRLGNIDIALMKLAGPLTFNEMIOVCLPNSSENFPPQKCVWTSWGATTELG 349
DB 641 IYINPHYKRRKRNNDIAMHMLEKMYNYDYIOPICLPENOVFPGRICISIAWGAL-1Y 699
OY 350 AGDASPVLNHAAPLISNKCINHR--DVGIIISPSMLCAGYLTGVDSCQDGSGLVYCO 408
DB 700 QGSTADVIGEADVPILSNKCCQOQMPHYN--ITENMVCAGYEAGVDSQDGSGLVYCO 757

OY 409 ERRLKLVGATSFQIGCAEVNKPQYTRVTSFLDWI 444
DB 758 ENRMVLGAVTSFGYOCALPNRPYARVPRTFEMI 793

RESULT 10

US-09-000-846-2
Sequence 2, Application US/09000846
Patent No. 5981830
GENERAL INFORMATION:
APPLICANT: WU, QINGYU
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
STREET: 2200 CLARENDON BLVD, SUITE 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000, 846
FILING DATE: 30-DEC-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/866, 058
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: LEBOVITZ, RICHARD M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: BERLX 65P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6410
TELEFAX: 703-243-6333
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 416 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-000-846-2

Query Match 26.8%, Score 655.5; DB 2; Length 416;
Best Local Similarity 38.0%; Pred. No. 4.6e-56;
Matches 143; Conservative 54; Mismatches 136; Indels 43; Gaps 10;

OY 100 DGEDEYRCVRYGGQNAVLYOF--TAASRKTWCSDDMKGHYANVACAQGFPSYSSDNLR 157
DB 46 DGEDEYRCVRYGGQNAVLYOF--TAASRKTWCSDDMKGHYANVACAQGFPSYSSDNLR 157
OY 158 VSLLEGQRFEEFVSIDHLLPDKVTALHSHYVREGSCAGHVVTLQCTAGH 209
DB 105 VRTAGANGTSFGYVDEBGLRLAQLL--DIVISV-----DCPRGRFLTATQDCGR 154
OY 210 RRGYSSRIVGNNMSLISQMPWQASIQFGYHLCGGSVITPLWITTAHCVYDLYLPR-- 266
DB 155 RKLPRYRIVGQDSSSLGHPMPQVSLRYDGTILCGSSLSDGMLVLAHC-----FPERNR 209
OY 267 ---SWTIOVGLVSLDNPAPSHLYEKIYHSHYKRRK---LGNIDIALMKLAGPLTFN 317
DB 210 VLSRWRVFAAGAVARTSPHAYVOLGYOVAVIYHGQYLPFRPTIDENSNDIALVLSLPLT 269
OY 318 EMIOVCLPNSSENFPPQKCVWTSWGATTELGAGDASPVLNHAAPLISNKCINHRDYYG 377

Db 270 EYIQVPLPAAGALVDKRYCTVIGWNTQ-FYGOAMVLQEARVPLISNEVCNSPDFYQ 328
QY 378 GLISPSMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGV 433
Db 329 NQKPMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGIYSWGTGICALARKPGV 388
QY 434 YTRVTSFLDWIHEOME 449
Db 389 YTKVTDREWIIFAIRK 404

RESULT 11

US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TADG-15
US-09-027-337-3

Query Match 23.5%; Score 576; DB 2; Length 256;
Best Local Similarity 43.1%; Pred. No. 1.4e-48;
Matches 114; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 216 RIYGGNMSLLSQMPWQASLOFQGYHLCGGSVITPLMTITTAHCYVDLYLPR-----SWT 269
Db 1 RIYGGNDTSLGRMPWQVSLRYDGAHLGCGSLSGDWLTAHC-----PERRNRVLSRMR 55
QY 270 IQVGLSLDNPAPSHL---VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFNEMT 320
Db 56 VFAGAVA---QASPHGLQGVQAVVYHGVLPRDPNSENSNDIALVHLSPLPLEYI 112
QY 321 QPCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 380
Db 113 QPCLPAGALVDKRYCTVIGWNTQ-YYGQAGVLAQEARVPLISNDVCGADFYGNQI 171
QY 381 PSMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGVYTR 436
Db 172 KPMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGIYSWGTGICALARKPGVYTR 231
QY 437 VNSFLDWIHEOME 449
Db 232 VSDREWIIFAIRK 244

RESULT 12

US-08-944-483-67
; Sequence 67, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:

INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6232456e

US-08-944-483-67

Query Match 23.3%; Score 571; DB 4; Length 255;
Best Local Similarity 44.8%; Pred. No. 4.4e-48;
Matches 113; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 217 IYGGNMSLLSQMPWQASLOFQGYHLCGGSVITPLMTITTAHCYVDLYLPR-----SWT 270
Db 1 IYGGNDTSLGRMPWQVSLRYDGAHLGCGSLSGDWLTAHC-----PERRNRVLSRMR 55
QY 271 IQVGLSLDNPAPSHL---VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFNEMT 321
Db 56 FAGAVA---QASPHGLQGVQAVVYHGVLPRDPNSENSNDIALVHLSPLPLEYI 112
QY 322 PCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 381
Db 113 PCLPAGALVDKRYCTVIGWNTQ-YYGQAGVLAQEARVPLISNDVCGADFYGNQI 171
QY 382 PSMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGVYTR 437
Db 172 KPMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGIYSWGTGICALARKPGVYTR 231
QY 438 TSFLDWIHEOME 449
Db 232 SDRREWIIFAIRK 243

RESULT 13

US-08-681-151-3
; Sequence 3, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael

APPLICANT: Goli, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0074US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 205011
US-08-681-151-3

Query Match 23.3%; Score 570; DB 2; Length 638;
Best Local Similarity 40.8%; Pred. No. 2,1e-47;
Matches 116; Conservative 52; Mismatches 88; Indels 28; Gaps 9;

QY 189 YVREGCASHVTL-----OCTACGHRGYSRIYGNMSLSQMPQASLOPQ---G 238
DB 362 YEAG-SSGYSLRCKVSESSDCT-----KINARIYGTNSLSGMPQVSLQVLSQ 415
QY 239 YHLCGGSVTPMLITTAHCYVDLYPKSWTQVGLSL--LDNPAHSLVEKIYHSKY 296
DB 416 NHMCGSIIIGRWILTAHCFDGIPIPDWRIYGLILNSETINKPFESSIKELIHQKY 475
QY 297 KPRRLGNDIALMKLGPLTFNEMIOVCLPENSENPPDKVQWTSQMGATEGAGASPV 356
DB 476 KMSGSDYDALIKLQPLTYTFQKPICLPSKADINTITYNCWYTWGVTKE--RGETONI 534
QY 357 LNHAAPLLSNRIC--NHRDVGIIISPSMLCAGYLTGVDSCQSGGSPVLCQOEIRLMK 414
DB 535 LKATIPLPVNECCOKKYDY---VITKMICAGYKEGIDACKGSGGSPVLCYKHSGRNQ 591
QY 415 LVGATSEFGICAEVKNPGVYTRVTSFLDWIHEOM---ERDIKT 454
DB 592 LVGITSMEGCAKKEQPGVYTVKVAEYIDWILEKIQSKERALET 635

RESULT 14
US-08-508-448C-25
Sequence 25, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:

APPLICANT: Kazuyoshi YAMAKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-448C-25

Query Match 23.1%; Score 564.5; DB 1; Length 418;
Best Local Similarity 34.2%; Pred. No. 3.9e-47;
Matches 142; Conservative 68; Mismatches 166; Indels 39; Gaps 13;

QY 51 IIVIGIALILALATIGLGHFDC--SGKYRCRSFEC--IELIARDGYSDDCKDGEDEYR 106
DB 22 IYVAGV--TLAVTIALVFLAFDQKSYFYRSSFOLLNVEYNOLN-----SPATQEYR 74
QY 107 CYRVGGONAVLOVFTASMKTM-----CSDMKGHYANVACA--OLGPPSYVSSDNLK 157
DB 75 TISGRIESLITKTFKESNLRNQFIRAHVAKLRDGSQVRAVDYMKQFTRNNGASMKSR 134
QY 158 VSLGQFREFEVSIDHLPPDKVTALHSHV---YVREGCASHVTLQCTACGHRREY 213
DB 135 IESVLRQMLNNSGNLE--INPSEITISLTQAAANMLINEGACPDILITS----- 183
QY 214 SSRIVGNNMSLSQMPQASLOFQGHLCGGSYTPMLITTAHCYVDLYPKSWTQVGL 273
DB 184 EORILGTFEAEBGSMPOVSLRLNNAHCGGSLINMMWILTAHCRSNSNPDMWATG 243
QY 274 LVSLDNPAPSHLVEKIYVHSKYKPRRLGNDIALMKLAGPLTFNEMIOVCLPENSENPP 333
DB 244 ISTTF--PKLRMRVNRILIHNNKSAFTHEMDIALVRLNSVFTKDIHSHVCLPAATONIP 301
QY 334 DGVKWTSGMGATGEGADASPVLNHAAPVLSNKCINHRDVGIIISPSMLCAGYLTG 393
DB 302 PGSTAVVTGMA--QEYAGHTVPELRQGVRIISNDVCNAPHYNGAILSGMLCAGVPDGG 360
QY 394 VDSQGDSCGPLYGQD--RLMLKLVGATSEFGICAEVKNPGVYTRVTSFLDWIHEQ 447
DB 361 VDACODSGPLVQEDSRILMFWIVGIVSWGDCGLPDKDGVYTVRYAVYIDWIRIQ 415

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds

(without alignments)
1251.566 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447
Sequence: 1 MGENDPPAVEAPFSPFSLFG.....TRVTSFLDWTHEQMERDLKT 454

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2447	100.0	454	1	TMS3_HUMAN
2	865.5	35.4	490	1	TMS2_MOUSE
3	864.5	35.3	492	1	TMS2_HUMAN
4	704	28.8	455	1	TMS5_MOUSE
5	666	28.4	457	1	TMS5_HUMAN
6	664.5	27.2	417	1	HEPS_HUMAN
7	661.5	27.0	1069	1	ENTK_MOUSE
8	660.5	27.0	437	1	TMS4_HUMAN
9	660	27.0	1035	1	ENTK_BOVIN
10	657	26.8	1019	1	ENTK_HUMAN
11	654	26.7	1034	1	ENTR_PIG
12	650	26.6	416	1	HEPS_MOUSE
13	645	26.4	416	1	HEPS_RAT
14	570	23.3	638	1	KAL_RAT
15	561	22.9	638	1	KAL_MOUSE
16	540	22.1	638	1	KAL_HUMAN
17	533	21.8	855	1	ST14_HUMAN
18	532	21.7	324	1	TEST_MOUSE
19	526	21.5	790	1	PLMN_PIG
20	523.5	21.4	761	1	NETR_MOUSE
21	520.5	21.3	338	1	PLMN_HORSE
22	520	21.3	1042	1	COR1_HUMAN
23	512.5	20.9	855	1	ST14_MOUSE
24	508.5	20.8	314	1	TEST_HUMAN
25	507.5	20.7	812	1	PLMN_MOUSE
26	507	20.7	437	1	ACRO_RAT
27	506	20.7	1113	1	COR1_MOUSE
28	504.5	20.6	810	1	PLMN_HUMAN
29	503.5	20.6	273	1	ACR7_MOUSE
30	503.5	20.6	436	1	ACRO_MOUSE
31	503	20.6	343	1	PLMN_SHEEP
32	501.5	20.5	810	1	PLMN_SHEEP
33	501.5	20.5	875	1	NETR_HUMAN

34	500.5	20.5	275	1	TRV1_CANFA	P15944 canis fam1
35	500.5	20.5	421	1	ACRO_HUMAN	P10323 homo sapien
36	499.5	20.4	342	1	PSS8_RAT	O96887 rattus norv
37	498.5	20.4	431	1	ACRO_RABIT	P48038 oryctolagus
38	498.5	20.4	812	1	PLMN_BOVIN	P06868 bos taurus
39	497.5	20.3	810	1	PLMN_MACRO	P12345 macaca mula
40	495.5	20.2	270	1	TRV1_MERUN	P50342 meriones un
41	495.5	20.2	333	1	PLMN_CANFA	P80009 canis fam1
42	495.5	20.2	342	1	PSS8_MOUSE	O96887 rattus norv
43	494.5	20.2	625	1	FAIL_HUMAN	P03951 mus musculu
44	490.5	20.0	276	1	MC16_MOUSE	P21845 mus musculu
45	488.5	20.0	274	1	TRV1_RAT	P50343 rattus norv

ALIGNMENTS

RESULT	ID	TMS3_HUMAN	STANDARD	PRT	454 AA.
AC	P57727				
DT	20-AUG-2001	(Rel. 40, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE				
DE	TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).				
GN	TPRSS3 OR TADG12 OR ECHOS1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RA	SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).				
RC	TISSUE=Ovarian carcinoma; PubMed=11068177;				
RA	MEDLINE=20521358; PubMed=11068177;				
RA	Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,				
RA	Wang Y., Parmley T.H., O'Brien T.J.;				
RT	"Ovarian tumor cells express a novel multi-domain cell surface serine				
RT	protease";				
RL	Biochim. Biophys. Acta 1502:337-350(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D), AND VARIANT ILE-53.				
RX	MEDLINE=20578749; PubMed=11137999;				
RA	Scott H.S., Rudoh J., Wattenhofer M., Shibuya K., Berry A., Chast R.,				
RA	Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,				
RA	Youniss F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C.,				
RA	Rossier C., Korostilshvsky M., Gal A., Shimizu N., Bonne-Tamir B.,				
RA	Antonarakis S.E.;				
RT	"Insertion of beta-satellite repeats identifies a transmembrane				
RT	protease causing both congenital and childhood onset autosomal				
RT	recessive deafness";				
RL	Nat. Genet. 27:59-63(2001).				
CC	- FUNCTION: PROBABLE PROTEASE.				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND				
CC	TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED				
CC	ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.				
CC	- DISEASE: DEFECTS IN TPRESS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL				
CC	NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 LDC-RECEPTOR CLASS A DOMAIN.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				

DR EMBL: AF201380; AAG37012.1; -
 DR EMBL: AB038157; BAB20077.1; -
 DR EMBL: AB038158; BAB20078.1; -
 DR EMBL: AB038159; BAB20079.1; -
 DR EMBL: AB038160; BAB20080.1; -
 DR MIM: 605511; -
 DR MIM: 6051072; -
 DR MIM: 605316; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001234; Trypsin.
 DR Pfam: PF00057; Idl_recept_a; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00192; Idla; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; tryp_Spc; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS50068; LDLRA_2; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KM Hydroxylase; Serine protease; Transmembrane; Signal-anchor; Deafness;
 Alternative splicing; Polymorphism.
 FT DOMAIN 1 48
 FT TRANSMEM 49 69
 FT DOMAIN 70 454
 FT DOMAIN 72 108
 FT DOMAIN 109 205
 FT DOMAIN 217 454
 FT ACT_SITE 257 257
 FT ACT_SITE 304 304
 FT ACT_SITE 401 401
 FT SITE 216 217
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT DISULFID 92 107
 FT DISULFID 129 194
 FT DISULFID 142 204
 FT DISULFID 207 324
 FT DISULFID 242 258
 FT DISULFID 338 407
 FT DISULFID 370 386
 FT DISULFID 397 425
 FT CARBOHYD 221 221
 FT VARSPPLIC 1 127
 FT VARSPPLIC 318 454
 FT VARSPPLIC 261 293
 FT VARSPPLIC 294 454
 FT VARIANT 53 53
 FT CONFLICT 46 54
 FT CONFLICT 90 90
 FT CONFLICT 350 350
 FT CONFLICT 369 395
 FT CONFLICT 427 427
 FT CONFLICT 454 AA; 49404 MM; 57ECC3678FD6A6F CRC64;
 SO SEQUENCE

Query Match 100.0%; Score 2447; DB 1; Length 454;
 Best Local Similarity 100.0%; Pred. No. 6; 9e-201;
 Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGENDPRAVEAPSFERSLFGIDLDKISPVAPDADAVAAQILSLPLKFFPIIYIGIALL 60
 Db 1 MGENDPRAVEAPSFERSLFGIDLDKISPVAPDADAVAAQILSLPLKFFPIIYIGIALL 60
 QY 61 LALAIGHIFHDCSGKYRCRSSFEKIELIARCDGVSDCKDGEDEYRCVRVGGNAVLQV 120
 Db 61 LALAIGHIFHDCSGKYRCRSSFEKIELIARCDGVSDCKDGEDEYRCVRVGGNAVLQV 120
 QY 121 TAASWTKMCDMDKGHANANACQOLGPPSYSSDNLRVSSLEQFREFPISIDHLLPDDK 180
 Db 121 TAASWTKMCDMDKGHANANACQOLGPPSYSSDNLRVSSLEQFREFPISIDHLLPDDK 180
 QY 181 VTALHSHVYREGCASHVVTLOCTACGHRGYSRRIVGGMNLSLSPMPQASLOFGYH 240
 Db 181 VTALHSHVYREGCASHVVTLOCTACGHRGYSRRIVGGMNLSLSPMPQASLOFGYH 240
 QY 241 LCGSVITPMTITTAHCYVDLYLPKSMTOVGLVSLDNPAPSHLYEKIVYHSKYPKR 300
 Db 241 LCGSVITPMTITTAHCYVDLYLPKSMTOVGLVSLDNPAPSHLYEKIVYHSKYPKR 300
 QY 301 LGNDIALMKLAGPLTFENMIQPCLPNSENPFDPKCYKWTSGMGATEDGAGDASPVLNHA 360
 Db 301 LGNDIALMKLAGPLTFENMIQPCLPNSENPFDPKCYKWTSGMGATEDGAGDASPVLNHA 360
 QY 361 AVPLISKICNHRDVGIIISPSMLCAGYLTGGVDSQGGSGGPLVQOERLKLVGATS 420
 Db 361 AVPLISKICNHRDVGIIISPSMLCAGYLTGGVDSQGGSGGPLVQOERLKLVGATS 420
 QY 421 FGIGCAEVNRPGYTVRTSFLDWIHEQMERDLKT 454
 Db 421 FGIGCAEVNRPGYTVRTSFLDWIHEQMERDLKT 454
 RESULT 2
 ID TMS2_MOUSE STANDARD; PRT; 490 AA.
 TM TMS2_MOUSE
 AC Q9JIO8; Q9JIC4; Q9JY82;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
 DE TRANSMEMBRANE PROTEIN X).
 GN TMPRS52.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=BALB/C;
 RX MEDLINE=21104370; PubMed=11169526;
 RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikio P.T.;
 RT "Expression of transmembrane serine protease TMPRS52 in mouse and
 human tissues";
 RL J. Pathol. 193:134-140(2001).
 RP [2]
 RP SEQUENCE FROM N.A.
 RA Han J., Kim S.;
 RT "Putative transmembrane protease X";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
 RT "A novel mosaic serine protease, epitheliasin";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

[illegible][illegible]


```

FT DOMAIN 71 455 (POTENTIAL).
FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
OX SRCR.
FT DOMAIN 218 455 SERINE PROTEASE.
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 217 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 14 MISSING (IN ISOFORM 2).
FT VARSPLIC 10 10 MISSING (IN ISOFORM 3).
FT VARSPLIC 182 182 MISSING (IN ISOFORM 1).
FT VARSPLIC 183 192 GGLVEAMKP -> MEAOVGLILMV (IN ISOFORM 1).
FT CONFLICT 325 325 D -> G (IN REF 1; BAB20277).
SQ SEQUENCE 455 AA; 49632 MM; 5FCF31789C6899AA CRC64;

```

```

Query Match 28.8%; Score 704; DB 1; Length 455;
Best Local Similarity 35.1%; Pred. No. 1.8e-52;
Matches 150; Conservative 63; Mismatches 162; Indels 52; Gaps 11;

```

```

OY 52 IVIGIALIALAIG--LGIHEDSGKRYRCSSFFCIELIARCDSVSDCKDEDEYR-- 106
DB 50 VILGVLGLLAGAGIASWMLVLYLMPAAPSISCTIQEEMTLNCPGV-C---EEELIDS 105
OY 107 -----CVRRGGONAVYQVTTAA--SKRTWCSDMKGHYANVACAQIGFSPYSVSDNLRVS 159
DB 106 LPTVSFRINGEDLLQOVYRARPDLVLCHEGSPALCMHICKSLGHLRLTQHKAVNLIS 165
OY 160 SLEQREFEVSIDHLIPDDKYATLHNSVYVREG-----CASGHVVTLOCTAC 207
DB 166 DIKLNSSQERFAQS-----ARRGGLVEAMKPSAACPGRIVSLKSCSC 209
OY 208 GHRGYSSTRIVGNSMILLSQWEPWASLQFQYHLCGGSVITPLMTTAAHCYVDLYLPR- 266
DB 210 G-ARPLASRIVGQAVASRWPQASVMLGSRHSCASVLAHMYVTAACHMSFRLSRL 268
OY 267 -SWTIQVGLVSLDNPAPSH---LVKEIYHSHKYPKRGNDIALMKLAGPLTFNEMIQP 322
DB 269 SSRRVYAGLVS--HGAVRQHGQTMWEKILPHPLYSQNHNDYDALLQRTPLNFSPTVDA 326
OY 323 VCIIPNSEENFPDQKVCWTSGWATEDGADASPVLNHAAPLISNKCINHRDVGGLIIP 382
DB 327 VCLPAKEQYFPMGSCQWGWCHTDPSSHSSDTLODTWVPLLSLTLCSCKSSCMGALNH 386
OY 383 SMCAGYLTVGVDSCGDSGGRLVCOERLMLKATSGICAEVYKPCGYTRVTSFLD 442
DB 387 RMLCAGYLDGRADACGDSGGRLVCPGSDTWMHVGVSWSGRCABEFPNRYAKVAEFLD 446
OY 443 WHEOME 449
DB 447 WIDHTVQ 453

```

```

RESULT 5
TMS5_HUMAN ID TMS5_HUMAN STANDARD; PRT: 457 AA.
AC 09H3S3;
DT 20-AUG-2001 (rel. 40; Created)
DT 20-AUG-2001 (rel. 40; Last sequence update)
DE TRANSMEMBRANE PROTEASE, SERINE 5 (EC 3.4.21.-) (SPINESIN).
GN TMRPSS5.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of human spinosin."
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AB028140; BAB20375.1; -.
CC InterPro: IPR001190; SRCR.
CC InterPro: IPR001254; Trypsin.
CC InterPro: IPR001314; Chymotrypsin.
CC Pfam: PF000089; trypsin.1.
CC PRINTS: PRO0722; CHYMOTRYPSIN.
CC SMART: SM00202; SR; 1.
CC SMART: SM00202; TRYP-Spc; 1.
CC PROSITE: PS00240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; FALSE_NEG.
CC PROSITE: PS00420; SRCR_1; FALSE_NEG.
CC PROSITE: PS00287; SRCR_2; FALSE_NEG.
CC KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor;
CC Glycoprotein.
CC
CC DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT DOMAIN 71 457 (POTENTIAL).
CC FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
CC FT ACT_SITE 218 457 SERINE PROTEASE.
CC FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT SITE 217 218 CLEAVAGE (POTENTIAL).
CC FT DISULFID 135 196 BY SIMILARITY.
CC FT DISULFID 148 206 BY SIMILARITY.
CC FT DISULFID 209 328 BY SIMILARITY.
CC FT DISULFID 243 259 BY SIMILARITY.
CC FT DISULFID 374 390 BY SIMILARITY.
CC FT DISULFID 401 429 BY SIMILARITY.
CC FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 457 AA; 49574 MM; 64406BA4953A2651 CRC64;

```

```

Query Match 28.4%; Score 696; DB 1; Length 457;
Best Local Similarity 35.2%; Pred. No. 8.6e-52;
Matches 149; Conservative 58; Mismatches 176; Indels 40; Gaps 7;

```

```

OY 53 VIGITALIALAIGLGIHEDSGKRYRCSSFFCIELIARCDSVSDCKDEDEYRCVRYVG 112
DB 51 VILGVLGLLAGAGVGSMLV---LYLCPAASQPIIS-----GLQDEEITLSCSEASA 98
OY 113 QNAVY-----QVFTASWKTKMCSDDMKGHYANVACAQIGFSPYS 152
DB 99 EEAALLPALPKTVSRINSEDFLENAQVRDPRMLVLCHEGSPALGQICWSLGLRLTH 158

```



```

RESULT 7
ENTR MOUSE
ID ENTR MOUSE STANDARD: PRT: 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRS57 OR ENTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Ductum;
RX MEDLINE=98147142; PubMed=946188;
RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinase (enteropeptidase) and expression in
RT small intestine during development.";
RT Am. J. Physiol. 274.G342-G349(1998).
CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
CC TRYPSINOGEN.
CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
CC SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U73378; AAB37317.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -.
DR MGD: MGI:1197523; PRS57.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000998; MAM.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLa; 2.
DR SMART: SM00137; MAM; 1.
DR SMART: SM00200; SEA; 1.

```

```

DR SMART: SM00202; SR; 1.
DR SMART: SM00020; TRYP_Spc; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLa_1; 2.
DR PROSITE: PS01209; LDLa_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00740; MAM_2; 1.
DR PROSITE: PS50060; MAM_2; 1.
DR PROSITE: PS50024; SEA; 1.
DR PROSITE: PS50020; SRCR_1; FALSE_NEG.
DR PROSITE: PS50020; SRCR_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HTS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KW Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 1 829
FT CHAIN 2 1069
FT TRANSMEM 19 47
FT DOMAIN 52 169
FT DOMAIN 227 268
FT DOMAIN 270 379
FT DOMAIN 387 549
FT DOMAIN 569 679
FT DOMAIN 686 724
FT DOMAIN 723 816
FT DOMAIN 830 1069
FT ACT_SITE 874 874
FT ACT_SITE 925 925
FT ACT_SITE 1021 1021
FT ACT_SITE 1021 1021
FT LIPID 2 2
FT DISULFID 229 242
FT DISULFID 236 255
FT DISULFID 249 266
FT DISULFID 688 700
FT DISULFID 695 713
FT DISULFID 707 722
FT DISULFID 817 845
FT DISULFID 859 875
FT DISULFID 959 1027
FT DISULFID 991 1006
FT DISULFID 1017 1045
FT CARBOHYD 147 147
FT CARBOHYD 197 197
FT CARBOHYD 212 212
FT CARBOHYD 373 373
FT CARBOHYD 380 380
FT CARBOHYD 433 433
FT CARBOHYD 515 515
FT CARBOHYD 579 579
FT CARBOHYD 675 675
FT CARBOHYD 727 727
FT CARBOHYD 751 751
FT CARBOHYD 770 770
FT CARBOHYD 791 791
FT CARBOHYD 897 897
FT CARBOHYD 936 936
FT CARBOHYD 999 999
SQ SEQUENCE 1069 AA; 118735 MW; E62549E463743C3D CRC64;

```

```

Query Match 27.0%; Score 661.5; DB 1; Length 1069;
Best Local Similarity 37.6%; Pred. No. 26-48;
Matches 152; Conservativity 64; Mismatches 143; Indels 45; Gaps 16;
OY 67 LGIHFDC-SGKYRCSSFCIELIARCDSVSDCKDEDEYRCVYR---VGQNAVQLQVFT 121
DB 682 LGIPEPCQDDEFQCKDQK-NCIPLGNLCDSYPCRDSDSACVRFNNGTRNSNGLOVFN 740
OY 122 AASWKTMCSDDMKGHYANVACAGLGFPSYSSNINLAVSSILEGFRREFPYSIHLLPDDKY 181
DB 741 HSIWHTACAEWMTQTSINEVCHLLIGLS--ANSMPISSTGG--GPFVRVND----- 788
OY 182 TALHSHVYVREG--CASGHVVTLOCT--ACGHR---GYSSRIYVGNMISLISQWPAQSL 234

```

```

Db 789 -APNGSLILPNSQSDSLILLCQCNKSCGCKVKTKVSPKIVGSGDAQAGAPRWVAL 847
OY 235 QFGCYH-----LCGSVITPMTITAHACYDYL-ERSWITQVGL--VSLIDNP-A 282
Db 848 ----YHNRDSTDRLLCGASLVSSDMLVSAHCVRRNLDPRMTAVGLHMOSLSTPQV 903
OY 283 PSHIVERTVYHSKKPKRLNDIALMKLAGPLFENEMIOPVCLPNSSENPDDGKVCWTSG 342
Db 904 VRRVVDIVINPHDRKRRKNDIAMHLEKRVNTDIOICLPENQIITPGTCIAG 963
OY 343 WGATEDGAGDASPVLNAAVPLISNKTICNHR-DVYGGIISPSMLCAGYLGTGVDSGCGS 401
Db 964 WGYDKINAGSTVDPLKEADVDPLISNEKCOQOLPEYN--ITESHICAGYEGEGIDSCGDS 1021
OY 402 GGPLVCOERLKLKVGATSPGICAEVNGKGVYTRVTSFLDWMH 445
Db 1022 GGPLMCEENNRWFLVGTSGVQCALPNHPGVYRVSOFTLEWH 1065

RESULT 8
TMS4_HUMAN STANDARD: PRT; 437 AA.
ID TMS4_HUMAN STANDARD: PRT; 437 AA.
AC Q9NR54; Q9NR54; 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 4 (EC 3.4.21.-) (MEMBRANE-TYPE SERINE
DE PROTEASE 2) (MT-SP2).
GN TMPRSS4 OR TMPRSS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Pancreatic carcinoma;
RX MEDLINE-20283276; PubMed-10825129;
RA Walltrap C., Haehnel S., Mueller-Pillaach F., Burghardt B.,
RA Iwanura T., Rubenbueger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMPRSS3) overexpressed in
RT pancreatic cancer.";
RL Cancer Res. 60:2602-2606(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Llanevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
CC FORMATION AND TUMOR INVASION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
CC AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
CC GASTROINTESTINAL AND UROGENITAL TRACT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF179224; AAF74526.1; -.
CC EMBL; AF216312; AAF31436.1; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR002172; LDL_receptor_A.

```

```

DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00057; ldl_recept_a.1.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDla.1.
DR SMART: SM00202; SR.1.
DR SMART: SM00202; TRYP-Spc.1.
DR PROSITE: PS01209; LDLa.1; FALSE_NEG.
DR PROSITE: PS00668; LDLa.2; FALSE_NEG.
DR PROSITE: PS00420; SRCR.1; FALSE_NEG.
DR PROSITE: PS50287; SRCR.2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM.1.
DR PROSITE: PS00134; TRYPSIN_HIS.1.
DR PROSITE: PS00135; TRYPSIN_SER.1.
KW Hydrolyase; Serine protease; Transmembrane; signal-anchor.
FT DOMAIN 1 38
FT TRANSMEM 1 38
FT SIGNAL-ANCHOR (POTENTIAL).
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 60 437
FT DOMAIN 61 93
FT DOMAIN 94 194
FT SRCR.
FT ACT_SITE 205 437
FT ACT_SITE 245 245
FT ACT_SITE 290 290
FT ACT_SITE 387 387
FT SITE 204 205
FT DISULFID 64 83
FT DISULFID 77 92
FT DISULFID 127 183
FT DISULFID 140 193
FT DISULFID 196 310
FT DISULFID 230 246
FT DISULFID 356 372
FT DISULFID 383 410
FT CARBOHYD 130 178
FT CARBOHYD 178 178
FT CONFLICT 1 31
SQ SEQUENCE 437 AA; 48204 MW; 351B2FD4A8557B12 CRC64;

Query Match 27.0%; Score 660.5; DB 1; Length 437;
Best Local Similarity 35.4%; Pred. No. 8.6e-49;
Matches 157; Conservative 74; Mismatches 149; Indels 63; Gaps 17;

OY 44 LPLKFFIYIGTALLALA-----IGGHPDGGKRRKSSPFC--IELIAR--- 91
Db 24 IPMEFFKVGIPITILLASIIIVVLKIKYLD--KY-----YELCGPLHFIPIPKOL 76
OY 92 CDGVSDEKDEDEYRC-----VRVGONAVLQVTTAA--SWKTMCSDDMKGHYA 138
Db 77 CDGGLDPLDEDEHCKYSFPEGPAVAVRLSKRSTLYQVDSATGNMFSACPFNFTALA 136
OY 139 NVACAQLGPPSYVSSDNLRVSSLEGQFEEFVSDILLPDKVTALHHSYVREG---CA 195
Db 137 ETACRQWGYSS-----KPTFRAYEIGIPDDOLDVVEITENSQELRMNRSSGQCL 184
OY 196 SGHVTYTCACGHRGYSRIYGGNMSILSOMPQASLOFGYHLCGGSVITPMTITA 255
Db 185 SGSLVSLHCLACG-KSLKTRVAVGGEASVDMWQVSIQYDQKHGCGSILDPHWHVLA 243
OY 256 AHCV--YDLTPKSWTIQVGLSLDNPAPSHIVERTV--YHSKKPKRLNDIALMK 309
Db 244 AHCRKRTDVF---NMKVRAGSDKL--GSPSLAVAKIIIEENPMV--PK--DNDAIALMK 255
OY 310 LAGPLFENEMIOPVCLPNSSENPDDGKVCWTSGMGATEDGAGDASPVLNAAVPLISNKT 369
Db 296 LQPLTFSGTVRPICLPFDEELTPATPLMIQWGTGFKONGKMSDILLQASVOVIDSTR 355
OY 370 CNHRDVGYGIIISPMICAGYLGTGVDSGCGDGGPLVCOERLKLKVGATSPGICAEVN 429
Db 356 CNMDDAVQGEVTERKMCAGIPEGGVDTCQDSSGSLPMTYSQD--MHVAVGIYSWGYCGGSP 414

```


FT	CARBOHYD	179	179	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD <td>328<td>328<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	328 <td>328<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	328 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>335<td>335<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	335 <td>335<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	335 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>388<td>388<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	388 <td>388<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	388 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>440<td>440<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	440 <td>440<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	440 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>470<td>470<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	470 <td>470<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	470 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>503<td>503<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	503 <td>503<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	503 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>534<td>534<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	534 <td>534<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	534 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>630<td>630<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	630 <td>630<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	630 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>682<td>682<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	682 <td>682<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	682 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>706<td>706<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	706 <td>706<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	706 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>725<td>725<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	725 <td>725<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	725 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>848<td>848<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	848 <td>848<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	848 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>887<td>887<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	887 <td>887<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	887 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>909<td>909<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	909 <td>909<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	909 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD <td>949<td>949<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td></td>	949 <td>949<td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td></td>	949 <td>N-LINKED (GLCNAC. . .)<th>(POTENTIAL)</th></td>	N-LINKED (GLCNAC. . .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CONFLICT <td>134<td>134<td>O -> E (IN REF. 3)<td></td></td></td></td>	134 <td>134<td>O -> E (IN REF. 3)<td></td></td></td>	134 <td>O -> E (IN REF. 3)<td></td></td>	O -> E (IN REF. 3) <td></td>	
FT	CONFLICT <td>732<td>732<td>S -> P (IN REF. 3)<td></td></td></td></td>	732 <td>732<td>S -> P (IN REF. 3)<td></td></td></td>	732 <td>S -> P (IN REF. 3)<td></td></td>	S -> P (IN REF. 3) <td></td>	
FT	CONFLICT <td>754<td>771<td>SOCLOSLIRLQCNHKS -> RRNAKEIDALSPILIA (IN REF. 3)<td></td></td></td></td>	754 <td>771<td>SOCLOSLIRLQCNHKS -> RRNAKEIDALSPILIA (IN REF. 3)<td></td></td></td>	771 <td>SOCLOSLIRLQCNHKS -> RRNAKEIDALSPILIA (IN REF. 3)<td></td></td>	SOCLOSLIRLQCNHKS -> RRNAKEIDALSPILIA (IN REF. 3) <td></td>	
SO	SEQUENCE	1019 AA:	112923 MW:	B6AA245FD4A563 CRC64:	
Query Match					
Best Local Similarity		26.8%;	Score 657;	DB 1;	Length 1019;
Matches 148;		Conservative 148;	Pred. No. 4.6e-48;	Indels 30; Gaps 15;	
QY	LGIHEDC-SGKYRCSRSKCIETLARCQGVSDCKDGEDEYRCVR- - - VGGQNAVLOYFT	121	121	121	121
DB	LGIEPPCCADHFQCKNG-ECYPLVNLCDGHLCEGDSADCVRFNCTTNNGLVPRRI	695	695	695	695
QY	122 AASWTKMSDWMKGHYANVACAOIGFPSSVDNLRVSSLEGQFNEFPVSIDHLLPDDKV	181	181	181	181
DB	696 QSIWHTACAEWNTQIISNDVQCLLGISGNSK- - -PIRSTDG- - -GPVKL-NTAPDG- -	747	747	747	747
QY	182 TALHSYVYREGCASGHVYVLTQCT--ACGHR- - -RGYSRIYVGMSLLSQMPQASIQF	236	236	236	236
DB	748 - - -HLILPSSQOCLQDSILIRLQCNHKSCKGKLLAADDITPKIVGGSNAEGAMPVAVGLY	804	804	804	804
QY	237 QGYHLGGSVTPPLMIIITAAHCYVDLYL-PSKWTIOVL- - -VSLIDNP-ASHSLVEKITV	292	292	292	292
DB	805 GGRLLCGSLVSSDMVSAHCVGRNLEPSPKATILGLHKSNTISQYTPRLIDELVI	864	864	864	864
QY	293 HSKTKPKRLGNDIALMKLAGPLTENEIMIQPVCLPNSSENFDPDGKVCWTSGKATIEDGAGD	352	352	352	352
DB	865 NPHNRRRKDDIAMHMLFEFVNTDYTOPICLPEENOVFPFGRCNSJAGWG- - -TVVYOGT	923	923	923	923
QY	353 ASPVLNHAAPVILISKICINHR-DVYGGIISPMCAAGLTVGVSOCGSDSGPLVCOERR	411	411	411	411
DB	924 TANIIQEAADVLLSNERCQOQMPEN- - -ITENMICAGTEBGGIDSCGSDSGPLMCOENN	981	981	981	981
QY	412 LMKLVGATSEFGICAEVKKPGVYTRVTSFLDWI	444	444	444	444
DB	982 RMFLAGVTSFEGYKCALPBRPGVAVARSRFTEMI	1014	1014	1014	1014
RESULT 11					
ID	ENTK_PIG	STANDARD;	PRT;	1034 AA.	
AC	P98074;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).				
GN	PR57 OR ENTK.				
OS	Sus scrofa (Pig).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.				
OX	NCBI_TaxID=9823;				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Duoanal mucosa;				
XC	MEDLINE=94327548; PubMed=8051081;				
XX					

QY 351 GASPVLNHAAPLISNKCINHR-DVYGGIISPSMLCAGLTGGVDSGCGSGPLVCOE 409
 DB 937 GSPADILGADVPLISNKCQOQMPYNN-ITENMACAGEEGCGDSCGSLMCL 994
 QY 410 RLRLKLVGATSFSGICAEVKKPGVYTRVTSFLDWT 444
 DB 995 NNRMLAGVTSFGYOCALPRRPGVAVRPKFTEMI 1029

RESULT 12
 HEP5_MOUSE STANDARD: PRT: 416 AA.
 AC 035453;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=98058912; PubMed=9395459;
 RA Yu T.-H., Liu R.-W., Haakma C., Tomasek J.J., Howard E.W.;
 RT Identification and cloning of the membrane-associated serine
 protease, hepsin, from mouse preimplantation embryos.
 RL J. Biol. Chem. 272:31315-31320(1997).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: AF030065; AAB84221.1;
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; -.
 DR MGD: MGI:1196620; Hpn.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001190; SSCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00202; SR.1.
 DR SMART: SM00020; TRYP-SPEC.1.
 DR PROSITE: PS00240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS.1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
 FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 (POTENTIAL).
 FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
 FT ACT_SITE 162 416 SERINE PROTEASE.
 FT ACT_SITE 202 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 256 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 352 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 187 203 BY SIMILARITY.
 FT DISULFID 321 337 BY SIMILARITY.
 FT DISULFID 348 380 BY SIMILARITY.

FT CARBOHYD 111 111 N-LINKED (GLCNAC...)(POTENTIAL).
 SQ SEQUENCE 416 AA; 44739 MW; 432194FF4004FB48 CRC64;
 Query Match 26.6%; Score 650; DB 1; Length 416;
 Best Local Similarity 38.5%; Pred. No. 6,3e-48;
 Matches 136; Conservative 49; Mismatches 128; Indels 40; Gaps 8;

QY 121 TASWKTMCSDDKGHYANACAGFPSPVSSDNLRYSLSEQFPREFPSID----- 173
 DB 68 TEGTWRLCSSRSNANAVAGLGECEMGLRALASELDVRAGANGTSGFPCVDEGLPLA 127
 QY 174 -HLPPDKVYALHHSYVREGCAGHVYTLQCTACGHRGSSRIYGGNMSLSQWPMQA 232
 DB 128 QRL- DVSVC-----DCPRGRFLATCCDGRKLPVRIYGGDSSLGRRPMQY 177
 QY 233 SLQFGYHLGGSVITPLMITTAACVYDLYDK-----SWTIQYGLVSLDNPAPSHL 286
 DB 178 SLRYDGTHLGGSLSGDWVLTAAHC-----FPERNRVLSRMHVPFAGAVARTSPHAVQLG 232
 QY 287 VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFENMTOPVCLPNSSENPFGKVCWT 340
 DB 233 VQAVYTHGGTLPDPPTIDENSNIDIALVHLSSPLREYIQVCLPAAQALVDGAVCYV 292
 QY 341 SGWATEDGADGASPVLNHAAPLISNKCINHRDVYGGIISPSMLCAGLTGGVDSGCGD 400
 DB 293 TGGNGNQ-FYGGQAMVLOEAPVPIISNVCNSDFGNGDKPMKPCAGREGIDACQGD 351
 QY 401 SGGPLVCOE---RLRLKLVGATSFSGICAEVKKPGVYTRVTSFLDWT 449
 DB 352 SGGPFCEDSISGTSWRLCGIVSMCTGALARKPGVYTRVTFRMRIFKAIR 404

RESULT 13
 HEP5_RAT STANDARD: PRT: 416 AA.
 AC 005511;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=93305733; PubMed=8318546;
 RA Farley D., Raymond F., Nick H.;
 RT Cloning and sequence analysis of rat hepsin, a cell surface serine
 protease.
 RL Biochim. Biophys. Acta 1173:350-352(1993).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: X70900; CA50256.1; -.
 DR PIR: S32013; S32013.
 DR PIR: S33777; S33777.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; -.

DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM0020; TRYP_SPE; 1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
 KM CHAIN 1 161
 FT CHAIN 161
 FT DOMAIN 162 416
 FT TRANSMEM 17 43
 FT DOMAIN 44 416
 FT ACT_SITE 162 416
 FT ACT_SITE 202 202
 FT ACT_SITE 256 256
 FT ACT_SITE 352 352
 FT DISULFID 152 276
 FT DISULFID 187 203
 FT DISULFID 321 337
 FT DISULFID 348 380
 FT CARBOHYD 111 111
 SQ SEQUENCE 416 AA; 44926 MW; E5A9F8FA9550E180 CRC64;

Query Match 26.4%; Score 645; DB 1; Length 416;
 Best Local Similarity 35.2%; Pred. No. 1.7e-47;
 Matches 146; Conservative 61; Mismatches 156; Indels 52; Gaps 11;

OY 60 ILALAIGLIHFDSCSKYRCSFSCIELIARCDGVSDCKGDEYKCVAVGONNAVLOY 119
 DB 17 VALVYGTLLFLTGIG---AASMAVITILR-----SDQPLVQVQLSPGDSRLVL 65
 OY 120 -FTASMKWTVCSDMKGHANVACAOLEFPYSVSDNLVYSSLEGQFREESVID----- 173
 DB 66 DKETGWRLLCSSRSNARVAGLGCCEGFLALAHSELVDRTAGANGTSGFVDEGLP 125
 OY 174 ---HLRDKVTALMHSYVREGCASGHVYTLDTCTACGHRGSSRYVGNMLSQMPW 230
 DB 126 LAQRL--DVISVC-----DCKRGRFLATCQDCGRKRLVDRTVGGDSSSLGWPW 175
 OY 231 QASLQFQGYHLCCGSVYTPMLITTAACVYDLYPK-----SWTIQVGLVSLDNPAPS 284
 DB 176 QVSLARDGTHLCCGSLSGMWVLTAAHC-----FPERNRVLSRRMYFAGAVATSPHAYQ 230
 OY 285 HIVEKIVYHSKYRPKR-----LGNDIALMKLAGPLTFENMIQPVCLPNSSENFDPDGKVC 338
 DB 231 LGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSSLPFLTEYIQPVCLPAAGQALVDGKVC 290
 OY 339 WTSCKGATEDGAGDASPVLNHAANVPLISNKGIRHRYVGGIISPSMLCAGYLGVGDSQC 398
 DB 291 TVTGMKNTQ-FYGGQAVVLOEARVPLISNVCSPDYQKIKPKMFCAYPGEGIDACQ 349
 OY 399 GDSGGLVCOER-----RLMKLVGATSEFGICAEVNRPGVYTRVTSFLDWIHEOME 449
 DB 350 GDSGGLVCOER-----RLMKLVGATSEFGICAEVNRPGVYTRVTSFLDWIHEOME 449
 RESULT 14
 KAL_RAT STANDARD: PRT; 638 AA.
 AC P14372;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-NOV-2001 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN PK.
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129236; PubMed=1993180;
 RA Beaudien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
 RT "Gene structure and chromosomal localization of plasma kallikrein.";
 RL Biochemistry 30:1628-1635(1991).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=90091743; PubMed=2598771;
 RA Seidah N.G., Lidenheim R., Mbikay M., Hamelin J., Lutfalla G.,
 RA Rougeon F., Lazure C., Chretien M.;
 RT "The cDNA structure of rat plasma kallikrein.";
 RL DNA 8:563-574(1989).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XII, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 CC EMBL; M62357; AAA74563.1; -
 CC EMBL; M62358; AAA74563.1; JOINED.
 CC EMBL; M62346; AAA74563.1; JOINED.
 CC EMBL; M62347; AAA74563.1; JOINED.
 CC EMBL; M62349; AAA74563.1; JOINED.
 CC EMBL; M62350; AAA74563.1; JOINED.
 CC EMBL; M62351; AAA74563.1; JOINED.
 CC EMBL; M62352; AAA74563.1; JOINED.
 CC EMBL; M62353; AAA74563.1; JOINED.
 CC EMBL; M62354; AAA74563.1; JOINED.
 CC EMBL; M62355; AAA74563.1; JOINED.
 CC EMBL; M62356; AAA74563.1; JOINED.
 CC EMBL; M30282; AAA41463.1; -
 CC EMBL; M58590; AAA42069.1; -
 CC PIR; A39180; KORTPL.
 CC HSSP; P00750; IRTF.
 CC MEROPS; S01.212; -
 CC InterPro: IPR000177; Apple.
 CC InterPro: IPR001314; Chymotrypsin.
 CC InterPro: IPR003014; PAN.
 CC InterPro: IPR001254; Trypsin.
 CC Pfam; PF00024; PAN; 4.
 CC Pfam; PF00089; Trypsin; 1.
 CC PRINTS; PR00005; APPLEDOMAIN.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC SMART; SM00223; APPLE; 4.
 CC SMART; SM00202; TRYP_SPE; 1.
 CC PROSITE; PS00495; APPLE; 4.
 CC PROSITE; PS0240; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SER; 1.
 CC Hydroxylase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KM Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Repeat.
 FT SIGNAL 1 19

FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA: 71368 MW: CC27C93AB1086599 CRC64;

Query Match 22.9%; Score 561; DB 1; Length 638;

Best Local Similarity 34.8%; Pred. No. 4e-40;

Matches 126; Conservative 61; Mismatches 109; Indels 66; Gaps 12;

OY 132 DMKGHANACAOGLFPFVSVDNLVSSLEGGFREFEVSIDHLPPDDKVTALHHSYVR 191
DB 301 DEGEELNV-----TEVOGADVCEETCTKTRCOFF-IYSLPDD-----CKE 342
OY 192 EGC-----ASGHVYTL-----QCTACGHRGRYSRIYGGNMS 223
DB 343 EGCCKSLRLSTDGSPTRITYGMOGSSGYSLRLCKLVDSPDCT-----KINARIYGGTNA 397
OY 224 LLSQWFMQASLOFO---GYHLCGGSVITPLMIITAAHCYVDLYLPKSWTIQVGLVSL-L 278
DB 398 SIGEMFMQVSLQVKIVSQTHLCGSIIGROWVTAHCFDGIPIYDPVWRIYGGILSLSEI 457
OY 279 DNPAPSHLEKIVYHSKYPKRLGNDIALMLAGPLTFNEMIQVCLPENSENFPDGKVC 338
DB 458 TKETPSSRIKELIHQERYVSEGNVDIALIKLOTPLNTEFOKPLCLPSKADNTNTYTNC 517
OY 339 WTSGMGATEDGAGDASPVLNHAAPLISNKIC--NHRDVGIIISPSMLCAGYLTGVD 396
DB 518 WYTGWGYTKE-QGETQNIILQKATIPLPVNBECQKKRYD---VINKOMICAGYKEGTDA 573
OY 397 CCGDSGGPLVCOERRLMKLVGATSPGICAEVKKPGVYTRVYSFLDWIHQME----RDL 452
DB 574 CKGDSGGPLVCKHSGRMQLVGITSMWEGCGCRKDPGVYTKVSEYMDWILEKTQSSDVRAL 633
OY 453 KT 454
DB 634 ET 635

Search completed: April 11, 2002, 08:57:04
Job time: 217 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 25.01 Seconds
(without alignments)
1344.633 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447

Sequence: 1 MGENDPPAVEAPFSEFSLFG.....TRVNSFLDWIHEQMRLDKT 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2432.5	99.4	453	22	AAE06935
2	2430.5	99.3	453	20	ABA44250
3	2418	98.8	452	20	AAV41694
4	2375	97.1	454	21	AAV32246
5	1788	73.1	327	21	AAV72093
6	1354	55.3	294	21	AAV32247
7	1339	54.7	248	21	AAV3572
8	1034	42.3	193	22	AAV3966
9	876.5	35.8	492	21	AAV92050
10	876.5	35.8	492	22	AAV01315
11	875.5	35.8	492	21	AAV77726

12	875.5	35.8	492	21	AAV44406	Human 20P12-GTC2
13	868.5	35.5	393	22	AAV01317	PI000C partial ami
14	864.5	35.3	492	21	AAV36901	Human TMPS52 prot
15	864.5	35.3	492	21	AAV57280	Ovrl15 homolog pro
16	864.5	35.3	492	22	AAE06943	Human serine prote
17	864.5	35.3	492	22	AAE01294	Human transmembran
18	731	29.9	421	22	AAV85042	Human SER6 protein
19	731	29.9	421	22	AAE01943	Human transmembran
20	727	29.7	688	22	AAE01944	Human transmembran
21	704	28.8	445	21	AAV1698	Human serine prote
22	704	28.8	445	21	AAV08950	Human secreted pro
23	702	28.7	414	21	AAV08912	Human secreted pro
24	696	28.4	457	21	AAV1699	Human serine prote
25	686.5	28.1	283	21	AAV81492	Human prostate-ass.
26	667	27.3	423	21	AAE06944	Human transmembran
27	664.5	27.2	417	22	AAE06942	Human enterokinase
28	661	27.0	273	21	AAV1696	Mouse serine prote
29	660.5	27.0	435	22	AAV06437	Mouse serine prote
30	660.5	27.0	435	22	AAV72558	Human protease HUP
31	660	27.0	798	15	AAV57283	Human seripancrin
32	657.5	26.9	492	22	AAV72559	Bovine enterokinase
33	657	26.8	1019	22	AAE06940	Human seripancrin
34	656	26.8	311	21	AAV1697	Human enterokinase
35	655.5	26.8	416	20	AAV43325	Mouse serine prote
36	655.5	26.8	416	20	AAV96812	A mouse serine prote
37	652	26.6	432	21	AAV99417	Human PRO1570 (UNO
38	652	26.6	432	22	AAV87581	Human PRO1570 (UNO
39	652	26.6	432	22	AAV66166	Protein of the inv
40	648.5	26.5	437	22	AAE06931	Human membrane-ty
41	646	26.4	172	22	AAV25245	Human protein sequ
42	600.5	24.5	238	21	AAV1695	Mouse serine prote
43	583.5	23.8	317	22	AAV25633	Human protein sequ
44	576	23.5	256	22	AAV89501	Human hepsin serin
45	564.5	23.1	418	17	AAV89435	Trypsin-like enzym

ALIGNMENTS

RESULT 1	AAE06935	standard; Protein; 453 AA.
ID	AAE06935	
XX	AAE06935;	
AC	16-OCT-2001	(first entry)
XX		
DE	Human membrane-type serine protease (MTSP) 6.	
DE	Human membrane-type serine protease; membrane-type serine protease;	
DE	MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;	
KW	lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy.	
KW		
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	/label= Transmembrane_domain
FT	Domain	/label= LDL_receptor_domain
FT	Domain	/label= LDL_receptor_domain class a"
FT	Domain	/label= SR_domain
FT	Domain	/note= "Scavenger receptor Cys-rich domain"
FT	Domain	/note= "Trypsin-like serine protease domain"
FT	Domain	/note= "Protease_domain
FT	Misc-difference	324
FT	Misc-difference	/note= "Unpaired cysteine"
XX		
PN	WO200157194-A2.	

XX Sequence 453 AA;
SQ
Query Match 99.3%; Score 2430.5; DB 21; Length 453;
Best Local Similarity 99.6%; Pred. No. 6e-187;
Matches 452; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 1 MGENDPAAVAPRPSRSLGIDDLKTSVPAPDADAQAOLSLPLKFPFIYIGITALL 60
DB 1 mgendppaveapstfslrlgldlkispvpadadavaaqlslplkfpfiylgltall 60
QY 61 IALAIALGIIHFDSCGKYRCKSSFKCIELIARCGVSDCKDEDEYRCVRYGONAVLOVF 120
DB 61 lalaialgihfdscgkyrcssfkciellarcgvsdckdegeyrcvryggnavlvqvf 120
QY 121 TAAASMTKCSDDMKGHYANVACAQLGPPSYSSDNLRVSSLEGQFNEEYVIDHLLPDDK 180
DB 121 taaswktmcsddmkghyanvacagqlgfpsysdnlrvsslegqfneefvsihlllpddk 180
QY 181 VTLAHSVYVREGCASGHVYTLQCTACGHRGYSRRIVGNSMLSLQMPWQASLQFGYH 240
DB 181 vtlahsvyvregcasghvylqctacghrgrysrrivgnsmlslqmpwqaslgfgyyh 240
QY 241 LCGSVITPLMITTAACHVYDLYLPKSWTQVGLVSLDNPAPSHLYEKIVYHSKTKPKR 300
DB 241 lcgsvitplmittaachvdydlylpkswtqvglvslldnpapshlyekivysktpkr 300
QY 301 LGMDILMKIAGPLTFNEMIQPVCLPNSSENPDPGKVCWTSGMGATDGAGDASPVLNHA 360
DB 301 lgmdilmkiagpltfnemiqpvcclpnssenpdpgkvctwtsgmgatdgagdaspvlnha 360
QY 361 AVPLISNKNICNHRDYGCGIITSPSMLCAGYLTGVDSCQSGGSPYVCOEAEKKIKVATS 420
DB 361 avplisnknicnhrdyggilispmlcagyltgvdscqsggspylvcgerllkvlgats 419
QY 421 FGIGCAEVNKPgyYTRVTSFLDIHEQMERDLEK 454
DB 421 fgigcaevnkpgyytrvtstfldwheqmerdlkt 453
RESULT 3
ID AAY41694 standard; Protein; 452 AA.
XX AAY41694;
AC
XX
DT 07-DEC-1999 (first entry)
XX
XX Human PRO382 protein sequence.
DE
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
secreted protein; transmembrane protein.
XX
XX Homo sapiens.
OS
XX
XX
PN WO946281-A2.
XX
PD 16-SEP-1999.
XX
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 13-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.

PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.
 XX
 PA (GETH) GENENTECH INC.
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX
 DR WPI: 1999-551358/46.
 DR N-PSDB; AA233949.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 12; Fig 28; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA233891 to
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 452 AA:
 SQ

Query Match 98.8%; Score 2418; DB 20; Length 452;
 Best Local Similarity 99.6%; Pred. No. 6e-186; 0; Indels 2; Gaps 2;
 Matches 452; Conservative 0; Mismatches 0;

QY 1 MGENDPPAVAPSPFSRSLFGIDLKISPVADDAVAQAIIISLPLKFPPIVIGITALLI 60
 DB 1 mgendppaveapspfsrslfgldlkisvpaddavaqaallslplkffflivgitali 59
 QY 61 LALATLGLGHFDCSGKRCRSSFRCIEELARCDGVSCKDGEDEYRCVYRGONAVLOY 120
 DB 60 lalatlglghfcdsgkrcrssfrcielarcdgvsckdgedeyrcvrygqnavloy 119
 QY 121 TAAASWKTMCSDMKGHHYANACAOIGFPYSVSDNLRVSSLEGOFEFEYSIDHLLPDDK 180
 DB 120 taaswktmcddmkghyanavacaglgfpyssdnlrsvssleqgfreesidhllpddk 179
 QY 181 VTALHSHVYREGCASHVVTLOCTACGHRRGYSRRIYGNMSILSQPMQASLQFOGYH 240
 DB 180 vtalhshvyregcasghvvtlqctacghrrgysrriygnmsilseqpmqaslgfgyh 239
 QY 241 LCGGSVTTPTMTTAAACVVDLYLPKSWTQVGLVSLDDNPAPSHLYEKIVYHSKYPKR 300
 DB 240 lcggsvttptmttaaacvvdlylpkswtqvglvslldnpapshlyekivynhsykpk 299
 QY 301 LGGNDILMKIACPLTENEMIQPCLPNSSENFDPGKWCMTSGMGATEGAGDASPVLNHA 360
 DB 300 lggndilmkiaapltenemiqpclpnssefnfdpgkwcwtsmgatgedg_gdaspvlnha 358
 QY 361 AVPLISNKCINHRDVGIGIISPSMLCAGYLTGVVSCQDGSGBGLVCQERRLMKLVGATS 420
 DB 359 avplisnkcicnhrdvygigilspsmcagyltgvvscqsgsgplvcqerrlmlklygats 418
 QY 421 FGIGCAEVNKRKYVTRVTSFLDWIHQMERDLYT 454
 DB 419 fgigcaevnkrpyvtrvtsfldwihqmerdlyt 452

RESULT 4
 AAB32246

ID AAB32246 standard; Protein; 454 AA.
 XX
 AC AAB32246;
 XX
 DT 11-JAN-2001 (first entry)
 XX
 DE Tumour associated differentially-expressed gene 12 protein sequence.
 XX
 KW Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytototoxic human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
 XX
 OS Homo sapiens.
 XX
 PN WO200052044-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 02-MAR-2000; 2000WO-US05612.
 XX
 PR 03-MAR-1999; 99US-0261416.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'Brien TJ, Underwood LJ;
 XX
 DR WPI: 2000-533263/48.
 DR N-PSDB; AAA93842.
 XX
 PT DNA fragment encoding tumor associated differentially-expressed gene 12
 PT protein used for diagnosing and treating malignant hyperplasia and
 PT cancers including ovarian cancer -
 XX
 PS Claim 3; Figure 4; 118pp; English.

This invention relates to a novel transmembrane serine protease called
 CC tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
 CC located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249
 CC represent human TADG-12 cDNA and their corresponding protein sequences.
 CC A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
 CC product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
 CC exhibits cytosolic activity, and can be used in vaccines and in gene
 CC therapy. TADG-12 nucleotide and protein sequences are used in the
 CC diagnosis of malignant hyperplasia and cancers of the ovary, breast,
 CC lung, colon, prostate and other cancers where TADG-12 is overexpressed.
 CC TADG-12 is particularly used as tumour marker for early disease
 CC diagnosis. TADG12 proteins or fragments can be used to vaccinate an
 CC individual with cancer, suspected of having a cancer or at risk of
 CC getting cancer. Sequences AAA93846-A93853 represent PCR primers used for
 CC amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
 CC TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
 CC create anti-TADG-12 antibodies. Sequences AAB32251-B32269 represent
 CC TADG-12 peptides which target HLA, and may be used in a vaccine or for
 CC immune stimulation.
 XX
 XX Sequence 454 AA:

Query Match 97.1%; Score 2375; DB 21; Length 454;
 Best Local Similarity 97.6%; Pred. No. 1.7e-182;
 Matches 444; Conservative 2; Mismatches 7; Indels 2; Gaps 2;

QY 1 MGENDPPAVAPSPFSRSLFGIDLKISPVADDAVAQAIIISLPLKFPPIVIGITALLI 59
 DB 1 mgendppaveapspfsrslfgldlkisvpaddavaqaallslplkffflivgitali 60
 QY 60 ILALATLGLGHFDCSGKRCRSSFRCIEELARCDGVSCKDGEDEYRCVYRGONAVLOY 119
 DB 61 ilalatlglghfcdsgkrcrssfrcielarcdgvsckdgedeyrcvrygqnavloy 120
 QY 120 FTAASWKTMCSDMKGHHYANACAOIGFPYSVSDNLRVSSLEGOFEFEYSIDHLLPDD 179
 DB 121 ftaaswktmcddmkghyanavacaglgfpyssdnlrsvssleqgfreesidhllpdd 180

QY 180 KVTALHSHVYVREGASGHVYTLQCTACGHRGYSRIVGNNLSLSQPMQASLQFGY 239
 Db 181 kvetalhshvyrregasghvltlqctacghrgrysriviagnnslsypwqaslfqgy 240
 QY 240 HLCGGSVITPWTITTAHCVYDYLPRKSWTIQVGLVSLDNPAPSHLVEKIVYHSKTKR 239
 Db 241 hlcggsvitpwtitaahecydylprkswtiqlvslldnpapshlvekiyhsykykpk 300
 QY 300 RLGNDAIIMKLAGPLTFENEMIQPVCLPENSENFPDGKVCWTSGMGATPDGAGASPLNH 359
 Db 301 rlgndaiimklagpltfenemiqpvcipnseentpdgkvcwtsgwateqg-gdasplnh 359
 QY 360 AAVPLISNKNHRYDYGIIISPSMLCAGYLTGVDSCQDSCGPIVQCERRLMKLVGAT 419
 Db 360 aavplisnknhrdygygiiispsmlcagyltgvdscqdsqgpivcqerrlmklvgat 419
 QY 420 SFEGCAEVNKPQVYTRVTSFLDWIHQMERDLKT 454
 Db 420 sfegcaeavnkpqvtrvtsfldwlhqmerrdlkt 454

RESULT 5

AA72093 ID AAY72093 standard; Protein; 327 AA.

AC AAY72093;

DT 28-MAR-2001 (first entry)

DE Human serine protease #4 encoded by clone HCHAK72.

XX Human; serine protease; osteopathic; immunosuppressive; antiallergic;
 KW antiinflammatory; cytoskeletal; cardiac; neuroprotective; nootropic;
 KW neuroleptic; vulnerrary; ophthalmological; antibacterial; antiviral;
 KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
 KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
 KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
 KW systemic lupus erythematosus; male reproductive system disorder;
 KW testicular cancer; digestion and food absorption disorder; arrhythmia;
 KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
 KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
 KW cardiovascular disorder; ocular disorder; drug screening.

OS Homo sapiens.

PN WO200068247-A2.

PD 16-NOV-2000.

PF 05-MAY-2000; 2000WO-US12207.

PR 07-MAY-1999; 99US-0133239.

PR 20-MAY-1999; 99US-0135163.

PR 03-AUG-1999; 99US-0147005.

PR 09-SEP-1999; 99US-0152935.

PR 01-NOV-1999; 99US-0162979.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Shi Y, Young PE, Ni J;

XX WPI; 2000-679799/66.

XX N-PSDB; AAD02323.

XX New nucleic acid molecules encoding human serine protease polypeptides,
 PT useful for diagnosis, prevention and/or treatment of disorders e.g.
 PT osteoporosis, lupus erythematosus and Alzheimer's -
 XX
 PS Claim 12; Page 280-281; 289pp; English.
 XX
 CC The present sequence is human serine protease #4 from clone
 CC HCHAK72 (ATCC Deposit No: PTA27).

CC The invention relates to human serine proteases and their cDNA clones.
 CC It is used in methods for the diagnosis, prevention and treatment of
 CC various disorders related to serine protease such as bone formation
 CC disorders (osteoporosis), connective tissue disorders (arthritis),
 CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
 CC reproductive system disorders (testicular cancer), digestion and food
 CC absorption disorders (Crohn's disease), neurodegenerative diseases
 CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
 CC proliferative and cancerous conditions (acute myelogenous leukaemia),
 CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
 CC ocular disorders (glaucoma) and infectious diseases caused by bacteria,
 CC viruses, fungi or parasites. It is also useful for screening therapeutic
 CC compounds. Serine proteases are used as immunological probes or
 CC polymorphic markers for the identification of chromosomes, cells and
 CC tissues in biological samples, identification of male contraceptive
 CC agents, delivery of compositions to targeted cells expressing a
 CC receptor for serine protease, hybridisation probes and molecular weight
 CC markers. Serine protease nucleic acids are also useful in gene therapy.
 CC Note: The present sequence shown in page 280-281 of sequence listing has
 CC been assigned SEQ ID NO: 14. But the sequence, human serine protease
 CC epidermal growth factor (EGF)-like domain (AA72114) shown in
 CC page 12 is also referred as SEQ ID NO: 14.

SQ Sequence 327 AA;

Query Match 73.1%; Score 1788; DB 21; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.4e-135;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 MCSDDMKGHYANVACAQLGPPSYVSSDNLRVSSLGQFREPEVSIDHLDPDQVYALHHS 187
 Db 1 mcsddmkghyanvacagqlgfpsyvssdnlrvsslegfreefvsidhlppdkvtalhs 60
 QY 188 VYREGCASHSVYTLQCTACGHRGYSRIVGNNLSLSQPMQASLQFGYHLCGGSV 247
 Db 61 vyregcasghvltlqctacghrgrysriviagnnslsypwqaslfqgyhlcggsv 120
 QY 248 TPLWITTAHCVYDYLPRKSWTIQVGLVSLDNPAPSHLVEKIVYHSKTKRRLGNDAI 307
 Db 121 tplwitiahecydylprkswtiqlvslldnpapshlvekiyhsykykpkrlgndai 180
 QY 308 MKLAGPLTFENEMIQPVCLPENSENFPDGKVCWTSGMGATPDGAGASPLVNHAAVPLISN 367
 Db 181 mklagpltfenemiqpvcipnseentpdgkvcwtsgwateqgdsasplvnhavplism 240
 QY 368 KICNHRDYVGGIISPSMLCAGYLTGVDSCQDSCGPIVQCERRLMKLVGATSGFGCAE 427
 Db 241 kicnhrdygygiiispsmlcagyltgvdscqdsqgpivcqerrlmklvgatsfgigcae 300
 QY 428 VNKPGVYTRVTSFLDWIHQMERDLKT 454
 Db 301 vnkpqvtrvtsfldwlhqmerrdlkt 327

RESULT 6

AAB32247 ID AAB32247 standard; Protein; 294 AA.

AC AAB32247;

DT 11-JAN-2001 (first entry)

DE Tumour associated differentially-expressed gene 12 variant 3 protein.

KW Transmembrane serine protease; TABG-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytoskeletal; human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.

XX Homo sapiens.

XX WO200052044-A1.

XX

[illegible]

ID	Sequence	Score	Pred.	No. Mismatches	Indels	Gaps
08-FEB-2001 (first entry)	Human cancer associated protein sequence SEQ ID NO:1017.					
XX	Human; cancer associated gene; cancer antigen; detection; cancer;					
XX	diagnosis; cytoskeletal; proliferative; vulnerable; immunomodulator;					
XX	antidiabetic; antistaphylococcal; antihemorrhagic; antiarthritic; antiviral;					
XX	antiinflammatory; antithyroidic; analgesic; antibacterial; cardiant;					
XX	dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;					
XX	vasculoprotective; antipsoriatic; antiangiogenic; gene therapy; inflammation;					
XX	immune disorder; haematopoietic cell disorder; autoimmune disorder;					
XX	allergic reaction; graft versus host disease; organ rejection;					
XX	haemostatic; thrombolytic; cardiovascular disorder; infection;					
XX	neurological disease; drug screening.					
OS	Homo sapiens.					
PN	WO20005350-A1.					
XX	21-SEP-2000.					
PD	08-MAR-2000; 2000WO-US05862.					
PE	12-MAR-1999; 99US-0124270.					
PR	(HUMA-) HUMAN GENOME SCI INC.					
PA	Rosen CA, Ruben SM;					
PI	WPI: 2000-587533/55.					
XX	N-PSSB: AAC77781.					
DR	Novel isolated nucleic acids comprising sequences encoding peptides					
XX	useful for treating or diagnosing e.g. cancer -					
PT	Claim 11; Page 1598-1599; 2352pp; English.					
PS	AAC77607 to AAC78448 encode the human cancer associated proteins given					
XX	in AAB3338 to AAB44239. The proteins can have activities based on the					
CC	tissues and cells the genes are expressed in. Example of activities					
CC	include: cytotoxic; proliferative; vulnerrary; immunomodulator;					
CC	antidiabetic; antistaphylococcal; antihemorrhagic; antiarthritic;					
CC	antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;					
CC	dermatological; neuroprotective; cardiant; thrombolytic; coagulant;					
CC	nootropic; vasotrophic; antipsoriatic and antiangiogenic. The					
CC	polynucleotides and polypeptides can be used for preventing, treating or					
CC	ameliorating medical conditions and diagnosing pathological conditions.					
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from					
CC	the present invention may be used to treat immune disorders by activating					
CC	or inhibiting the proliferation, differentiation or mobilisation of					
CC	immune cells, to treat disorders of haematopoietic cells, autoimmune					
CC	disorders, allergic reactions, graft versus host disease and organ					
CC	rejection, modulate haemostatic or thrombolytic activity, modulate					
CC	inflammation, cancers, cardiovascular disorders, neurological disease and					
CC	bacterial or viral infections. The peptides, nucleosides, antibodies, and					
CC	agonists and antagonists may be also be used in drug screens. AAC78449 to					
CC	AAC78437 and AAB44240 represent sequences used in the exemplification of					
CC	the present invention.					
XX	Sequence 248 AA;					
SQ						
Query Match	54.7%; Score 1339; DB 21; Length 248;					
Best Local Similarity	100.0%; Pred. No. 1e-99;					
Matches 245:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	RRGYSRIVGVGMMSLLSQPMQASIQPGCYHLGGSVITPLMIITAAHCVDLYLPKSWT 269					
Dd	4 rrrysrrivggmmsllsqpmqasirgfyghlccgsavtlplmltaahcaydlylpkswt 63					
OY	IQVGLVSLINDNPADSHLEKIVHSKYRKRGNDIALMKTAGLPTFENMIQVCLPNSF 329					
Dd	iqvqvslindnpapshylekvhsrkyrkrgndialmktaglptlfemmqpvclpnse 123					

QY 330 ENPDKVCTSCMGATIEDGADSPVLAHAAPLISNKNICNHRDVGIIISPSMLCAGY 389
DB 124 enfpdkvctscmgatiedgagdasplvlnhaavplisnknichrdvgygllispsmlcagy 183
QY 390 LTGVDSGCGDSGGLVCOERRLMKLVGATSFEGICAEVKNPGVYRTVTSFLDWIHEOME 449
DB 184 ltgvdscgsgdsgglvvcgerllwklvgatsfgigcaevnkpvytrvtsfldwihgme 243
QY 450 RDLKT 454
DB 244 rdlkt 248

RESULT 8
AAG73966
ID AAG73966 standard; Protein; 193 AA.
AC AAG73966;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:4730.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
PD 05-APR-2001.
PF 28-SEP-2000; 2000WO-US26524.
XX
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH33397.
XX
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 6530-6531; 9803pp; English.
XX
XX
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing P.
Additionally, N may be used to produce the colon cancer-associated Ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AAB77789 represent sequences used in the exemplification of the
present invention.
CC
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX
SQ Sequence 193 AA;

Query Match 42.3%; Score 1034; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 2.2e-75;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PKSWTIOVGLVSLDNPAPBHLVEKIVYHSKYPKRLGNDIALMKLAGPLTFEMIOPCV 324
DB 4 pkswtiovgvlslndnpapbhlvekiyhskyprkrlgndialmklagpltfemioqpcv 63
QY 325 LPNSENFPDQKVCWTSYGWATEDGAGDASPVLAHAAPLISNKNICNHRDVGIIISPSM 384
DB 64 lpnsefnfpdqkvctscmgatiedgagdasplvlnhaavplisnknichrdvgygllispsm 123
QY 385 LCAGYLTGVDSGCGDSGGLVCOERRLMKLVGATSFEGICAEVKNPGVYRTVTSFLDWI 444
DB 124 lcagyltgvdscgsgdsgglvvcgerllwklvgatsfgigcaevnkpvytrvtsfldw1 183
QY 445 HEOMERDLKT 454
DB 184 heqmerdlkt 193

RESULT 9
AA92050
ID AA92050 standard; Protein; 492 AA.
AC AA92050;
XX
DT 01-AUG-2000 (first entry)
XX
DE HRPc6/7 polypeptide from androgen-inducible gene clone.
XX
KW Androgen inducible; testosterone; prostate cancer; cytostatic;
KW TMPRSS2; diagnosis.
XX
OS Homo sapiens.
XX
PN WO200018961-A2.
PD 06-APR-2000.
PF 30-SEP-1999; 99WO-US22535.
XX
XX
PR 30-SEP-1998; 98US-0163759.
PR 30-SEP-1998; 98US-0164159.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Macbeth KJ, Shyjan AW;
XX
DR WPI; 2000-293182/25.
DR N-PSDB; AAA08803.
XX
XX
PT Novel methods for identifying compounds for treating prostate cancer
PT comprising measuring the level of expression or activity of 1 or more
PT of 11 genes or their products
XX
PS Claim 2; Fig 3; 108pp; English.
XX
XX
This protein is encoded by a gene which is androgen (e.g. testosterone)
inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells)
and constitutively expressed in androgen-independent prostate cancer
cells (e.g. LNCap cells). Agents which decrease the expression or
activity of these clones may slow or arrest the growth of prostate cancer
cells or may kill them. HRPc6/7 can be obtained from the sequence of
the known gene for TMPRSS2. A compound useful for treating prostate
cancer can be identified in a novel method comprising measuring the
expression level, or activity, of HRPc2, 3, 6/7, 8, 9, 10, 13, 14, 15,
19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the
presence and absence of a test compound. The sequences may also be used
in diagnosis of prostate cancer and to determine efficacy of treatment
for prostate cancer.
CC
CC
XX
XX
SQ Sequence 492 AA;

RESULT 13
AAM01317
ID AAM01317 standard; Protein; 393 AA.
XX
AC AAM01317;
XX
DT 04-OCT-2001 (first entry)
XX
DE P1000C partial amino acid sequence 100-492.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
XX cytosolic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kales MD, Fanger GR, Day CH, Relter MW, Stolk JA, Skeiky YAM;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 540-541; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytosolic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (II) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 393 AA.

Query Match 35.5%; Score 868.5; DB 22; Length 393;
Best Local Similarity 46.6%; Pred. No. 9.7e-62;
Matches 187; Conservative 50; Mismatches 143; Indels 21; Gaps 11;
XX
OY 63 LAIGLGHF---DCSGK-TRCRSSFKCIELIARCDGVSDCKGDEDEYRCVRRGGNAVAUQ 118
XX
DB 1 laagl1wktmgscksgtscgclnpsnwcgvshncpgedencrvlrygsfllq 60
XX
OY 119 VFPA--ASWKTWCSDMKHGVANVACAOLGFP-SYSSDNLNLRSSLEGGFREFEVSIDHL 175
XX
DB 61 vvsqrkshpvcqddmnenygraactdmgyknnfytssgldvdsstsfmkntsaagnv 120
XX
OY 176 LPDDKVTALHHSVYREGCASHVVTLOCTACGRRGYS--SRIVCGNNSILSQWPMQAS 233
XX
DB 121 ---dlvkk1kylhs-----daccskavslrclacgvnlinsrgrstiyvgesalpgapwpyqs 173
XX
OY 234 LQFGYHLCGGSVITPLWITTAHCY-YDLVLPKSWTIOGVLV--SLDNPAPSHLVEKI 290

DB 174 lhvgnvhvcgssiltpewivlaahcvekp1nmgwhwtafagilrgsfmfyga-gyqvekv 232
OY 291 VYHSKXKPRRLGNDILMKIAGFLTFENEMQPCVLNSENSENFDPDGKVCWCMSCGATEDGA 350
DB 233 ishpnydstckkndialmk1qkpltdlvkpyclpnpqmm1lqpeqlcwisvgatlee-k 291
OY 351 GDASPVLNHAAPVLIENKICNHRDVGILISPSMLCAGYLTGVDSCQDGSQGPVLCOER 410
DB 292 gkxsevlinaakvllietqrnsryydnlltpmicagflqgnvdsqgdsqgplvtskn 351
OY 411 RLMKLVGATSFGICAEVKNPGVYTRVTSFLDIHQBOMERD 451
DB 352 nlwvlligdtswsgscakayrpygygvnmvftdwl1yrgmrad 392

RESULT 14
AAB36901
ID AAB36901 standard; Protein; 492 AA.
XX
AC AAB36901;
XX
DT 26-FEB-2001 (first entry)
XX
DE Human TMPRSS2 protein.
XX
KW Prostate specific androgen regulated protein; ARSRL; TMPRSS2;
KW PART-1; neoplastic.
XX
OS Homo sapiens.
XX
PN WO200065067-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10920.
XX
PR 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
PA (UNITV) UNITV WASHINGTON.
XX
PI Nelson PS, Hood L, Lin B;
XX
DR WPI; 2000-679676/66.
DR N-PSDB; AAC83325.
XX
PT Polynucleotide encoding prostate specific androgen regulated
PT polypeptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX
PS Claim 63; Page 86-88; 121pp; English.
XX

The present invention relates to prostate specific androgen regulated
CC proteins. The invention may be used to determine an expression level
CC of the prostate-specific proteins ARSRL; TMPRSS2, or PART-1 in a
CC fluid sample or prostate cell sample from an individual. It may also
CC be used for diagnosing and predicting the susceptibility of a
CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.
XX

Query Match 35.3%; Score 864.5; DB 21; Length 492;
Best Local Similarity 45.4%; Pred. No. 2.6e-61;
Matches 186; Conservative 55; Mismatches 146; Indels 21; Gaps 11;
OY 52 IVIGIITLALAIAGLGHF---DCSGK-TRCRSSFKCIELIARCDGVSDCKGDEDEYRC 107

Thu Apr 11 09:27:12 2002

us-09-846-512-12.rag

Page 12

Db 149 VRLGPHNFILOYVXSORKSWHPYQDDMWNENYGRACRDMGKKNFESSOGIIVDSGSTS 208
QY 165 FREEFSIDHLLPDDKVTALHSHVYVREGASGHVYVLTQCTAGCHRGYS--SRIYGNM 222
Db 209 FKLMTSAGNV--DIYKKLYHS---DACSSAAVSLRCLACGVNLNSSRQSIYGES 261
QY 223 SLISQMPQASLOFGYHLCGGSVITPLMTITTAHCV-YDLYLPKSWTIOVGLV--SLLD 279
Db 262 ALPGAMPQVSLHVNHHVHVGSGSIITPEWITVTAHCEKPLNPMWMTAFAGILRQSPMF 321
QY 280 NPAPSLHVEKIVYHSKYPKRLGNDIALMKLACPLTFNEMIQVCLPENSENPPDGKVCW 339
Db 322 YGA-GQVKEKIVSHPNYDSTKNNIDALMKLOKPLTFNEMIQVCLPENSENPPDGKVCW 380
QY 340 TSGMGATEGAGDASPVLNHAAPVLSNKNICNHRDYGGIISPSMLCAGYLTGVDSCG 399
Db 381 ISGMGATEE-KGKTEVYNAAKYLLETORCNSRYVDNLITPAMICAGFLQGNVSCG 439
QY 400 DSGGPLVCOERRLMKLVGATSFEGICAEVKNKPGVYTRVTSFLDMIHOMERD 451
Db 440 DSGGPLVTSKNNIMWLIGDTSMGSCAKAYRPGVGNVMTFTDWTYROMRAD 491
RESULT 2
Q9DGR2 PRELIMINARY; PRT; 767 AA.
ID Q9DGR2;
AC Q9DGR2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EMBRYONIC SERINE PROTEASE-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis";
RL Gene 252:209-216(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(LDLR) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL; AB038497; BAB08217.1; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam; PF00057; ldl_recept_a; 6.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLr; 8.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 33.7%; Score 825.5; DB 13; Length 767;
Best Local Similarity 45.0%; Pred. No. 5,7e-69;
Matches 175; Conservative 46; Mismatches 131; Indels 37; Gaps 10;

QY 78 RRRSFKETELIARCDVSDCKDEGEYRCVRVGNNAVLYQV--TAASWKTCSDMDKG 135
Db 394 RCGSSVSVCLSSQWCDGVSDCPYGEDEMGSCLYLPADFOLQVYSTSVSAMLPCVCSYDWMND 453

QY 136 HYANVACAQLGF--PSYVSDNLR-----VSLGQREEFVSDHLLPDDKVTALH 185
Db 454 DFRFRACQDFGIVGSSYNNRDTLMSYPANPGYFKLYSGWRKSF-----YTSVQ 502
QY 186 HSYVYREGCASGHVYVLTQCTAGC-HRRGYSSRIYGVGNMSLLSQMPQASLOFGYHLCG 244
Db 503 YSSY-----CVSGNVVSLHICSGVSNNSLVSRIYVGTFANLGMWPMQVNLQITGVLCG 558
QY 245 SVITPLMTITTAHCVYDLYLPKS-WTIQVGLVLLNP-----APSLHVEKIVYHSKYPK 299
Db 559 SLISPMVITAAHCVYGSYSASGAWRVFAG--TLTKPSYNSAAVEERIVHPGKSY 615
QY 300 RLGNDAIIMKLACPLTFNEMIQVCLPENSENPPDGKVCWTSGMGATEGAGDASPVLN 359
Db 616 TYNDIATLAKMLRDEITFGYITQVPCLPNSGMFEAGTTTWISMGSTYEG-GSVSTYLOY 674
QY 360 AAVPLISNKNICNHRDYGGIISPSMLCAGYLTGVDSCGDSGGPLVCOERRLMKLVGAT 419
Db 675 AAIPLIDSNVNCOSYVNGQITSSMICAGYLSGVDTCGDSGGPLVKNRNGTMTLVGDT 734
QY 420 SFGICAEVKNKPGVYTRVTSFLDMIHOM 448
Db 735 SWDGCARANKPGVGNVTFTEWITYSQM 763
RESULT 3
Q9BYE2 PRELIMINARY; PRT; 581 AA.
ID Q9BYE2;
AC Q9BYE2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MEMBRANE-TYPE MOSAIC SERINE PROTEASE.
GN MSPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmir S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung";
RL Biochim. Biophys. Acta 1518:204-209(2001).
DR EMBL; AB048796; BAB39741.1; -
KW Protease.
SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44A CRC64;

Query Match 29.8%; Score 728; DB 4; Length 581;
Best Local Similarity 37.9%; Pred. No. 6e-60;
Matches 157; Conservative 66; Mismatches 163; Indels 28; Gaps 9;

QY 44 LPLKFRPIIVIGIALILALIGLGHFDC-----SKYCRSSFKETELIARCDVSDCK 99
Db 161 LPL-----IGVLLILALVSLILILFQWGHTRIRKEQRESCPKAAVRCGVDDCK 213
QY 100 DGEDEYRCVYVGGONAVLVFTAA--WKTCSDMDKGHYANVACAQLGFPSVSSDNL 157
Db 214 LKSDELGCVRFMDKSLILYSSSHQWLPICSSNNNDSTSEKTCQLQDFESHRTTEVA 273
QY 158 VSLGQREEFVSDHLLPDDKVTALHSHVYVREGASGHVYVLTQCTAGHRRGYSSRI 217
Db 274 -----HRDFANSFSLRYN--STIQESLH-RSHCPQORYISLQCSHCQ-LRAWTGRI 321
QY 218 VGSNMSLLSQMPQASLOFGYHLCGGSVITPLMTITTAHCVYDL--YLPKSWTIOVGLV 275
Db 322 VGGALASDSKWPQVSLHFGTHIGGLTIDQWVITAAHCFVYREKYLEGKYYAGYS 381
QY 276 SLIDNPAPSLHVEKIVYHSKYPKRLGNDIALMKLACPLTFNEMIQVCLPENSENPPDG 335

```

Db      382 NH0LPEASIAE-IIINSNTYDEEDDYDIALMLRLSKPLTLIAHHPACLPMMGOTFSIN 440
QY      336 KYCWMTSGCATDGDGADSPVLNHAAPLISNKICNHRVYGGIISPMILCAGLYLGVD 395
Db      441 ETCWITGFGKTRDCKTSPLEVOVNLIDFKKCNLYVSYLLPRMKAGDLHGHD 500
QY      396 SCQSGSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOME 449
Db      501 SCQSGSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOME 554

```

RESULT 4

```

O9BEYL  PRELIMINARY:  PRT:  537 AA.
AC  O9BEYL:
DT  01-JUN-2001 (TREMBLrel. 17, Created)
DT  01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE  01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE  MOSAIC SERINE PROTEASE.
GN  MSPS.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=LUNG;
RX  MEDLINE=21167393; PubMed=11267681;
RA  Kim D.R., Sharmah S., Inoue M., Kido H.;
RT  "Cloning and expression of novel mosaic serine proteases with and
RT  without a transmembrane domain from human lung.";
RL  Biochim. Biophys. Acta 1518:204-209(2001).
DR  EMBL: AB048797; BAB39742.1; -.
KW  Protease.
SQ  SEQUENCE 537 AA; 58102 MW; A39F4E8B16DAECF CRC64;

```

Query Match 29.1%; Score 712.5; DB 4; Length 537;
 Best Local Similarity 38.9%; Pred. No. 1.5e-58;
 Matches 145; Conservative 64; Mismatches 147; Indels 17; Gaps 7;

```

QY      85 CIELIARCDVSDCKDGEDERCVRGVGNAYLOVFTAA--WKTWCSDMKGHVAVAC 142
Db      174 CRKHAVRCDGVVDCKLSDCLGCVREDMKSLKTIYSSGSHQMLPICSNNMDSYSEKTC 233
QY      143 AQLGPPSYSSONLNVSSLEGOFPREFVSIHLLPDDKTALHSHYVEGACSGHVTL 202
Db      234 ROLGFESAHRTTEVA-----HDFEANSFSLRYN--STIOESLH-RSHCPSORYISL 282
QY      203 OCTACGHRGYSRRIYVGNMISLQPMQASLOFOGYHLGCGSVITPLIITAHCYVDL 262
Db      283 QCSHSG-LRAMTGRIVGALASDKRPMOVSLHFGTHICGSLTLDQAVNLTAHCFEYT 341
QY      263 --YLPKSWTIQVGLSLDNPAPSHLVEKIYVSKYKPRKGLNDIALMLKLAGPLTFNEMI 320
Db      342 REKVLGEMKVYAGTSTNLHQLPEASIAE-IIINSNTYDEEDDYDIALMLRLSKPLTLIAH 400
QY      321 QVVCPLPNSSENPDDGKVCMTSGMGATEDGAGASPVLNHAAPLISNKICNHRVYGGI 380
Db      401 HPACLPMMHQOTFSLNETCITTFGKTRDCKTSPLEVOVNLIDFKKCNLYVSYLLPRMK 460
QY      381 SPSMLCAGLYTGVDSCQSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSF 440
Db      461 TTRMNCAGDLHGGRDSCQSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSF 520
QY      441 LPMIHEQMERDLK 453
Db      521 LPMIYSKMESEVR 533

```

RESULT 5
 097506

```

ID      097506  PRELIMINARY:  PRT:  643 AA.
AC  097506;
DT  01-MAY-1999 (TREMBLrel. 10, Created)
DT  01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT  01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE  KALLIKREIN.
OS  Sus scrofa (Pig).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX  NCBI_TaxID=9823;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=PLASMA;
RA  Takahashi T., Kimura A., Okimura H., Hamabata T.;
RT  "Porcine liver plasma kallikrein.";
RL  Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
CC  -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC  -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR  EMBL: AB022425; BAA37147.1; -.
DR  HSSP: P00766; 1CHG.
DR  MEROPS: S01.212; -.
DR  InterPro: IPR000177; Apple.
DR  InterPro: IPR001314; Chymotrypsin.
DR  InterPro: IPR003014; PAN.
DR  InterPro: IPR001254; Trypsin.
DR  Pfam: PF00024; PAN; 4.
DR  Pfam: PF00089; trypsin; 1.
DR  PRINTS: PR00005; APPLIEDOMAIN.
DR  PRINTS: PR00722; CHYMOTRYPSIN.
DR  SMART: SM00223; APPLE; 4.
DR  SMART: SM00020; TRYP-SPEC; 1.
DR  PROSITE: PS00495; APPLE; 4.
DR  PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR  PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
KW  Hydrolyase; Serine protease.
SQ  SEQUENCE 643 AA; 72227 MW; AFE2923E3C3CB80A CRC64;

```

Query Match 23.7%; Score 579; DB 6; Length 643;
 Best Local Similarity 36.9%; Pred. No. 7.4e-46;
 Matches 128; Conservative 55; Mismatches 118; Indels 46; Gaps 10;

```

QY      132 DMKGHYANACQLGPPSYSSONLNVSSLEGOFPREFVSIHLLPDDKTALHSHYVEGACSGHVTL 176
Db      309 DEGEELAN-----TFVGANLCOQECFTIRQFTYSLHPEDCGEKCKSLRIS 360
QY      177 PDDKYTALHSHYVVRG-----CASGHVVTLOCTACGHRGYSRRIYVGNMISLQPMQ 231
Db      361 SDGSPFKIRHGMKASSGYSRLCRSG---DHSACATKA--NTRIVGGTDSFLGEMPMQ 413
QY      232 ASDLPQ---GYHLGCGSVITPLIITAHCYVDLYLPKSWTIQVGL--VSLDNPAPSHL 286
Db      414 VSLQAKLRQNHLCGSSITIGHQVNLTAHCFDGLSLPDIWRIYGGIINISLTKETPESQ 473.
QY      287 VERIYHYSKPKRRLGNDIALMLKLAGPLTFNEMIQVCLPNSSENPDDGKVCMTSGMGAT 346
Db      474 VKEIITHQYKILLESCHDIALKLEPLVYTFQKRICLPNSRDTVVYTNCKVYTWGTF 533
QY      347 EGCAGDASPVLNHAAPLISNKIC--NHRDVGIGIISPMILCAGLYTGVDSCQSGDGP 404
Db      534 EE-KETIQNIIQLKVNPLVSNNECOKSYRD--HKISKMIAGYEGGKDACKEGSGP 589
QY      405 LYCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOMERD 451
Db      590 LYCKYNGIHLVGTTSWGECCARQEPGYTYIVIEYMDILEKTQDD 636

```

RESULT 6
 09DB10 PRELIMINARY: PRT: 799 AA.
 AC 09DB10;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 1300008A22RIK. PROTEIN.
GN 1300008A22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AK004939; BAB23684.1;
DR MGD: MG1:1919003; 1300008A22RIK.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001217; LDL_recept_A.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00057; ldl_recept_a; 3.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00192; LDLa; 3.
DR SMART: SM00020; tryp_Spec; 1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS50068; LDLa_2; 3.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Glycoprotein; Hydrolase; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A64A5D288 CRC64;

```

```

DB 558 LOGSSRIYGVTSSEGEPMWQASLQIRGRHICGALLIADRWVITTAHCFQEDSMASPKL 617
QY 268 WTQVGVGLNDLP----APSHLYEKIVYHSKPKPRRLGNDILMKLAGLPFNEMIOPV 323
DB 618 WTFLG--KMRQNSRWPEGVSEFVSRFLPHRYHEEDSHDYALDQDHPVYSATVRPV 675
QY 324 CLPNSSENPDDGKVCWTSGMGATEDGADASPVLNHAAPPLISNKNRHDVYGIISPS 383
DB 676 CLPARSHFEFGHCHMTGAGQREG--GPVSNLTQKVDVQVLPQDICS--EAVRYGVSPR 732
QY 384 MLCAGILTGVDSCQGDGSGPLVCOF--RLMKLVGATSRGICGAENVKRGVYTRVTFD 442
DB 733 MLCAGYRKGKKKACQDSDGSLVCRPSGRWFLAGLVSMGLGCGRPFGVYTRVTRVIN 792
QY 443 WHEE 446
DB 793 WIDQ 796

```

RESULT 7
09NUS5 PRELIMINARY: PRT: 1322 AA.

```

AC 09NUS5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE SERINE PROTEASE 22D.
GN SP22D.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxId=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3; TISSUE=HEMOLYMPH;
RX MEDLINE=20110889; PubMed=10646969;
RA Gorman M.J., Andreeva O.V., Paskewitz S.M.;
RT "Molecular characterization of five serine protease genes cloned from
RT Anopheles gambiae hemolymph.";
RL Insect Biochem. Mol. Biol. 30:35-46(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AF117751; AAD38337.3;
DR InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00530; SRCR; 2.
DR PRINTS: PR00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SMART: SM00494; CHBD2; 2.
DR SMART: SM00192; LDLa; 2.
DR SMART: SM00202; SR; 2.
DR PROSITE: PS01209; LDLa_1; 1.
DR PROSITE: PS50068; LDLa_2; 2.
DR PROSITE: PS00420; SRCR_1; UNKNOWN_1.
DR PROSITE: PS50287; SRCR_2; 2.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1322 AA; 146811 MW; 2707110783A6B843 CRC64;

```

Query Match 23.2%; Score 567; DB 5; Length 1322;

Best Local Similarity 31.3%; Pred. No. 2,6e-44;
Matches 144; Conservative 68; Mismatches 174; Indels 74; Gaps 15;

```

QY 53 VITIIALLALALGIGHDCSGKY-RCRSSFCIELIARCDVSPCKOEDEX--RC-- 107
Db 872 VGVVCRTPVMS-----CPDYLWLCCHASECIPVQFLCDVNRCCADSDSPHCKA 923
QY 108 ---VR-VGG---ONAVLQVFTASWKTMSDDMKGHYANVACAQLGFPSSVSDNLRVSS 160
Db 924 PLAVRLVAGTDEGSEVLEINHGWTGTCDDDDGCVREARICQLDGFNGIAEVRKSVYPP 993
QY 161 LBSQFREFEVSIDHLIP--DDKYTALHSHVYREGCASGHVTLQC----- 204
Db 984 GVGQIMLDVQACNGTEPSIEDCV---HMHGSEHNCANHEDEYVGCVVYTPKAPARLR 1039
QY 205 -----TAGC-----HRRGYSRTYVGNMSSLISQMPQASLQFO 237
Db 1040 ATRPNRPFDEVERSKIHDPDTCGRVLIDPTLRKPTYGARVHSGSEVYGHHPQASLRVK 1099
QY 238 GYHLCGGSVITPLMITTAACHVYDLYLPKS-WTIOVG--LVSLDNPAPSHLEKITYHS 294
Db 1100 TMMWCAVLITRHYVLTAAHCL--IGPKSTYVRIGDHTAAYDAELDIFIENTYIHE 1157
QY 295 KYRP-KRLGNDIALMKLAGPLTFNEMIQPVCLPENSENEPDKVYWTSGMATEDAGDA 353
Db 1158 QREGHMMSNDIAVVVLKTPVRENDYVOPICLPARDAPYLPQONCTISGSGATEAGSKDS 1217
QY 354 SPLNHAAPVLIJNKCINHDVYGGIISPSMCAGYLTGVDSCQDSDGSPVLC-QERRL 412
Db 1218 SYDLRAGVPLPDPSCRRPEYVGDSLIDMFCAGTLEPVDSCDDSDGSLVCPNSEGL 1277
QY 413 WKLVGATSEFGCAEYVKNPGVYTRVTSFLDMIHOMERDL 452
Db 1278 HTLTGIYSWKGKCGYANKPGVYLKVAHYRMIQKLNOSL 1317

RESULT 8
ID 060235 PRELIMINARY; PRT; 418 AA.
AC 060235;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AIRWAY TRYPSIN-LIKE PROTEASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98234382; PubMed=9565616;
RA Yamoka K., Masuda K., Ogawa H., Takagi K., Umemoto N., Yasuoka S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
like protease."
RT J. Biol. Chem. 273:11895-11901(1998).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AB002134; BAA28691.1;
DR HSSP: P00750; 1RTF.
DR MEROPS: S01.301;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR SMART: SM00200; SEA; 1.
DR SMART: SM00200; TRYP-Spc; 1.
DR PROSITE: PS00134; TRYPSIN HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 418 AA; 46263 MW; F4RCIDB020CFBBD0 CRC64;

```

Query Match 23.1%; Score .564.5; DB 4; Length 418;
Best Local Similarity 34.2%; Pred. No. 9.7e-45;
Matches 142; Conservative 68; Mismatches 166; Indels 39; Gaps 13;

```

QY 51 IIVIGIALLALALGIGHDC--SGKYRCSSFC--TELIARCDVSGDKGDEYR 106
Db 22 IVVAGVV--LVAVTALLVFLAFDQKSYFYRSFOLLANEVNSQLN-----SPATQEYR 74
QY 107 CVRVGQMANVLOVFTASWKTMS-----CSDMKGHYANVACA-QLGFPSSVSDNLR 157
Db 75 TLSGRIESLTKTFKESNLNNOFIKRAHVAKLRDQGSVRAADVWKKFQFTBNNGASMKSR 134
QY 158 VSLLEGQREFEVSIDHLIPDDKYTALHSHV---YVREGCAS-HVYTLQCTACGHRGY 213
Db 135 IESVLQMLNNSGNLE-INSTEITSLTDQAAANWMLINEGCAGPDLITLS----- 163
QY 214 SSRIVGNMSSLISQMPQASLQFOGYHLCGGSVITPLMITTAACHVYDLYLPKSWTIQVG 273
Db 184 EQRILGTEAEESGSPWQVSLRLNNAHHCSSGLINNMWILTAACHFRSNGNPRDWTATSG 243
QY 274 LVSLDNPAPSHLEKIVYHSKPKRLGNDIALMKLAGPLTFNEMIQPVCLPENSENEFP 333
Db 244 ISTTF--PKLRMRVRNILLNNYKSAETHENDIALVRLNENSVETPKDIHSVCLPAAQONIP 301
QY 334 DGKVCWTSGMATEDAGDASPVLNHAAPVLIJNKCINHDVYGGIISPSMCAGYLTG 393
Db 302 PGSTAYVTGGA-QETAGHTVPELRQGVKTIISNDVCNAPASHNGAILSLCAGVPQGG 360
QY 394 VDSQGDSDGSPVLCOE-RRLMKLVGATSEFGCAEYVKNPGVYTRVTSFLDMIHQ 447
Db 361 VDACQSGSGPLVQEDSRRLMFLVGIYSWGDQGLPDKPQVYTRVAYLDMIRQ 415

```

```

RESULT 9
ID 09NATO PRELIMINARY; PRT; 1322 AA.
AC 09NATO;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ADHESIVE SERINE PROTEASE.
GN SP22D.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20318993; PubMed=10860981;
RA Danilell A., Loukeris T., Lagueux M., Mueller H.M., Richman A.,
RA Kafatos F.C.;
RT "A modular chitin-binding protease associated with hemocytes and
RT hemolymph in the mosquito Anopheles gambiae."
RT Proc. Natl. Acad. Sci. U.S.A. 97:7136-7141(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AJ276428; CAB81934.1;
DR InterPro: IPR002557; Chitin binding.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001190; SCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.

```

DR SMART; SM00494; ChtBD2; 2.
 DR SMART; SM00192; LDla; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS01209; LDla_1; 1.
 DR PROSITE; PS0068; LDla_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_1.
 DR PROSITE; PS00287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Glycoprotein; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 1322 AA; 146794 MW; FBP973C21CC5475B CRC64;

Query Match 23.0%; Score 564; DB 5; Length 1322;
 Best Local Similarity 31.1%; Pred. No. 5, 1e-44;
 Matches 143; Conservative 68; Mismatches 175; Indels 74; Gaps 15;

QY 53 VIGIILALALGIGIHFDGSGKY-RCRSSFKCIELIARCDSVSDCKDEDEY--RC-- 107
 DB 872 VVGVCVCTPYMS-----CPDDYMLCHASSECIPOVFLCDNVRDCADGSDSPDHCKA 923
 QY 108 ---VR-VGG---QNAVQVFTASWKTMCSDMKGHYANACADLGFPSYSSDNLAVSS 160
 DB 924 PLAVRLVAGPTDREGREIVHYHGTWGTVCDDDFVREARVYICRQLGFNGTAEVAKSVYPP 983
 QY 161 LEGQFREEFASIDHLPR--DDKYALHHSYVREGASGHVYTLQC----- 204
 DB 984 GVGQIWLQVACNGTEPESIDCV---HHMHGHNCGHTEDVGRGCVYVPTKARARLR 1039
 QY 205 -----TACG-----HRRGYSSRIYVGMSLISQMPWQASLQFO 237
 DB 1040 ATBPMPFDEVERSRIKIHPTDSCGVLLDPLRLKPTYGARVHGHSEYVGHHPWQASLRK 1099
 QY 238 GYHLCGGSVTPMLITIAACGVVDLYLPRK--WTLQVG--LVSLDNPAPHVLYKVIYHS 294
 DB 1100 TMMHCGAVLITRHHVLAACHL--IGYPKSTYRIRIGDYHTPAAYDANLDFIENTYIHE 1157
 QY 295 KYRP-KRLGNDIALMKLAGPLTFENEMIQVCLPNSSENPFDGKVCWTSFGATEEDGAGA 353
 DB 1158 QFEGHMSNDIAVVLVKTQVRFNDYQPICLPARADAPYLPGQCTISGNGATEAGSKDS 1217
 QY 354 SPVLMHAAPVLISKICNHDVYGGIISPSMLCAGYLTGVDSCGDSGGRPLVC-QERRL 412
 DB 1218 SYDLRAQTVLPLDPSVCRREYVGDLSLDGMFCAGLEPVDSCDSDSGPLVCPSNSEL 1277
 QY 413 WKLVGATSFGICAEVKNPGVYTRVTSFLDMIHQEMERDL 452
 DB 1278 HTLTGIYSMGKHCYANKPGVYTLKVANHRTMIEKLNQSL 1317

RESULT 10
 ID 09Y1V3 PRELIMINARY; PRT; 868 AA.
 AC 09Y1V3;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DE TUNICATE RETINOIC ACID-INDUCIBLE MODULAR PROTEASE PRECURSOR.
 GN TRAMP.
 OS Polyandrocampa misakiensis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 OC Stolidobranchia; Styelidae; Polyandrocampa.
 OX NCBI_TaxID=7723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE SPOT.
 RA Ohashi M., Kawamura K., Fujii N., Yubisui T., Fujiwara S.;
 RT "A retinoic acid-inducible modular protease in budding ascidians.";
 RL Dev. Biol. 0:0-0(1999).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLRA) DOMAIN
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.

CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL; AB030007; BAA82522.1; -
 DR HSSP; P00750; 1RTE.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR003609; Pan_app.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00057; 1dl_recept_a; 3.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00192; LDla; 3.
 DR SMART; SM00473; PAN_AP; 1.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS01209; LDla_1; 3.
 DR PROSITE; PS0068; LDla_2; 3.
 DR PROSITE; PS00287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Glycoprotein; Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1
 SQ SEQUENCE 868 AA; 97659 MW; F71462865F36ACA CRC64;

Query Match 22.4%; Score 548.5; DB 5; Length 868;
 Best Local Similarity 32.3%; Pred. No. 8, 3e-43;
 Matches 146; Conservative 59; Mismatches 148; Indels 99; Gaps 19;

QY 72 DC--SGKYRRSPFKCIELIARCDSVSDCKDEDEYRC-----VRGQNAV-----LQVFT 121
 DB 434 DCATNTNYLNDG--SCLEHQQVCFRDPCNGEEMTECEPMKIFLRGSGPFEGHYEVAK 492
 QY 122 AASKWTMCSDMKGHYANVACAQLGF-----PSYSSDNLAVSSLEGO 164
 DB 493 GRKYGILCDTRMSTREADVVCRLQGFRRGAILDKLGAGFGRIDRRV--LDNVRNG---- 547
 QY 165 FREFEVSIDHLPRDCKTALHHSYVREGASGHVYTLQC----- 204
 DB 548 -RE--SLEN-----CRHSGMNAKASCSIDHEGVICRQETATPSPSPATPPTP 595
 QY 205 -----TACGHRGY-----SSRIYGMMSLISQMPWQASLQFOGYHLGGSVIT 248
 DB 596 RRTTTPPPMPBCGRKRVLEAPLPTARIYVGSSTGEHHPMQAGIVLPWTYMGCSGLIH 655
 QY 249 PLMTTAAHGVVDLYLPRKSMITQVG--LVSLDNPAPSHLVEKIYHVKYKPKRLGNDIA 306
 DB 656 PCWVLTAAHCFVREYPRIDYIRIGDHIITGVDETEOLFKAIEIKHD--YNVTTKENDIA 714
 QY 307 LMKLAGP----LTFENEMIQVCLPNSSENPFDGKVCWTSFGMATEEDGAGA--SPVLMHAA 361
 DB 715 LRLTENDARCATTPEVQVCLPKSSQFPDAKTTICEVTMGSDATVAAYVPLQEA 774
 QY 362 VPLISNKNCHRDVYGGIISPSMLCAGYLTGVDSCGDSGGRPLVCQERR-----LMLKV 416
 DB 775 IPLIANKK-LRSEVYQLQPTMFCAGYLTGVDSCGDSGGRPLVCQSDDDYVW--- 830
 QY 417 GATSFGICAEVKNPGVYTRVTSFLDMIHQEM 448
 DB 831 GIVSWGKCAKPKAPGVYAKAVAFIDWI--EOM 861

RESULT 11
 ID 09JUT7 PRELIMINARY; PRT; 855 AA.
 AC 09JUT7;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC


```

DE SERINE PROTEASE).
GN MBSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=JEJUNUM;
RA Tsuzuki S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=DUDENUM;
RA Inoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(CC (LDLR) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL: AB037898; BAB03502.1;
DR EMBL: AB049188; BAB13765.1;
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_recept_A.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDL; 3.
DR SMART: SM00020; tryp-spec; 1.
DR PROSITE: PS00339; AA_trna_ligase_II_2; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 4.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
FT VARIANT 665 K->N.
SQ SEQUENCE 855 AA; 94955 MW; 35806B7E6FC03D CRC64;
Query Match 22.2%; Score 542.5; DB 11; Length 855;
Best Local Similarity 31.5%; Pred. No. 3e-42;
Matches 135; Conservative 63; Mismatches 148; Indels 83; Gaps 16;
OY 73 CGGKRCRCSFICIELIACDGVSCDCKGDEDEYRCYRVGQNAVLQVFAAAMKRTMCSDD 132
DB 453 CGGMECKTG-CIKRDLNCDWADCPDSDERHC-----RCNATHQPCNKXQ 499
OY 133 -----WKGHYANVACAQLGFPYSVSDNLRVSLLEGQREFEVSIDHLLP----- 178
DB 500 FCKPLFWCDVSN-DCGD-----GSDEGSCSPAGSK-----CSNGKLLPQSQCKNGKD 548
OY 179 -----DKVTALHSHSVYRBCASGHVVTLLQCTACGHRGYS----- 214
DB 549 DCGDGSDEASCNVNAVSCCTKYTYR-CQNGLCINKNPCDCKKDCSDKNCDCGR 607
OY 215 -----SRVIGGMSLSLQWPMQASLQFGY-HLCGGSVITPLMITTAACVVDILPK-- 266
DB 608 SFTKQARVYGVGNADGEMPMQVSHALGQHLGASLSLPMVLVSAACFODETIFKYS 667
OY 267 ----SWTIQGLVSLDNPV----PSHLVERKIVYHSKYPKRLNDNLAMKALPLTFENMI 320
DB 668 DHTMTAFIAGLLDQSKRSASGVQEHKLRKRLITHPSNDTTPYDIALLEKPAEISTYV 727
OY 321 QPVCLPNSSENFPPDGKVCWTSGMGATEDGAGDASPVLAHAAPVLLISNKCINHRDVGII 380

```

```

DB 728 RPICLPDNTHPFAGKAIWTFGHTKEG-GTGALLIQKGEIIVINQTC--ELLPPQI 784
OY 381 SPSMLCAGYLTGCVSDCGDGGSPLYCOQR--RLMKLVGATSGICGACVNPQVYTRT 438
DB 785 TPRMKCVGLSGVSDSCQDGGSPVSSVERKRTFQ-AGVSWGEGCAORNNPQVYTRLP 843
OY 439 SFLDWIHEQ 447
DB 844 EYRDWIKED 852
RESULT 12
OYVSU2 PRELIMINARY; PRT: 1186 AA.
AC OYVSU2: OYVSU;
DT 01-MAR-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TEQUILA PROTEIN.
OS TEQUILA OR CG4821 OR CG4948 OR CG18403.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
NCBI_Taxid=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Ffannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jajami M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirker E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sviderskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(CC (LDLR) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
EMBL: AE003553; AAF50319.2;

```

DR HSSP; P20231; 1AAO.
 DR FlyBase; FBgn0023479; Tegulla.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR002965; P_rich_extensions.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF00057; ldl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; trypsin; 2.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR01217; PRICHEXTENSN.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYR_SPC; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein; Hydrolase; Serine protease.
 KW Glycoprotein; 1186 AA; 131246 MW; 75474D3968B8A32D CRC64;
 SQ SEQUENCE

Query Match 22.0%; Score 539; DB 5; Length 1186;
 Best Local Similarity 27.5%; Pred. No. 9.9e-42;
 Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

QY 66 GLGIH-----FDSCGKY-RCRSSFCIELIARCDGVDCKDEDEYRCV 108
 DB 709 GMEVHNGVDVAGYCKVYVMKCPNNYICHTSKECIPAFVCDNPPDADSDE--CA 766
 QY 109 RV-----GGQNA--VLOVFTAASKWTCSDMKGHYANVACAQLFPSSYSSDN 155
 DB 767 AVCOAPQYRLLEGGRNNEGRLVKKHGVGSCVDDDFNLSAQVACNSMGFEGPAKIEK 826
 QY 156 LRVSSLEGQREFEVSIDHL---PDDKYALHHSYVVRGCSAGHYVTLQCTA----- 206
 DB 827 NIFGNSNGP-----IWLDOVCFGENETSIDQCNHMMNGEHNCHTDEVALHCSAGPPRS 881
 QY 207 -----CG-----HRGYS 214
 DB 882 QRYSQIQIKGRSLGRFTTPKITYSQIGLMERSKAVHTPRRCGIFKDDLDTEVAHR---E 938
 QY 215 SRIYVGNMSSLQMPWQASIQFOG-----YHLCGGSVITPLMITTAHCVYDLYLPK-SW 268
 DB 939 ERYVVRGNVAQRGHRPWQATIRTRGRGSISSHMCAGAVVISKRHILTAHCLYG--SPKGA 966
 QY 269 TIQVG--LVSLLDNPASHLYEKIVHSKY-KKRLGNDIALMKLAGPLFNEMIQVCL 325
 DB 997 FTVVGDHYANIAISSSEVDFSTENMYLHENFRKGTTHMNDIALVVLKTPLEKFSVDQICL 1056
 QY 326 PNSEENEPDQKVCWTSGMGATEDEGADSPVLNHAAPVLSNKCINRDVYGGITPSML 385
 DB 1057 PDKNAELVEDRKCTISGWSIKSGVSTPRAVYLSAELPIILADHYCKOSNYYGSMSSGMF 1116
 QY 386 CAGYLTGGVDSQGDSCGGLVCOERRLMKLVGATSFSGICAEVNNKPGVYTVTSFLDMIH 445
 DB 1117 CAGSMESVYACBDSGSLVCSDDDEFTLYGLISWGHCGRKNNRPGVYVYVNHYIMY 1176
 QY 446 EOMERDL 452
 DB 1177 EKTINSL 1183

RESULT 13
 090112 PRELIMINARY; PRT; 1449 AA.

AC 090112;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GRAAL PROTEIN.
 GN TEQUILLA OR GRAAL OR CG4821 OR CG4948 OR CG18403.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mueller A.I., Medzhitov R., Janeway C., Hoffmann J.A., Laguex M.;
 RT "Characterization of a new serine protease in Drosophila
 RT melanogaster."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLRA) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL; A1251803; CAB64653.1; -.
 DR HSSP; P20231; 1AAO.
 DR FlyBase; FBgn0023479; Tegulla.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR002557; Chitin_binding.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF01607; Chitin_bind_2; 2.
 DR Pfam; PF00057; ldl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00494; ChlBD2; 2.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYR_SPC; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein; Hydrolase; Serine protease.
 KW Glycoprotein; 1449 AA; 160045 MW; F3BC806543CAB6ED CRC64;
 SQ SEQUENCE

Query Match 22.0%; Score 539; DB 5; Length 1449;
 Best Local Similarity 27.5%; Pred. No. 1.3e-41;
 Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

QY 66 GLGIH-----FDSCGKY-RCRSSFCIELIARCDGVDCKDEDEYRCV 108
 DB 972 GMEVHNGVDVAGYCKVYVMKCPNNYICHTSKECIPAFVCDNPPDADSDE--CA 1029
 QY 109 RV-----GGQNA--VLOVFTAASKWTCSDMKGHYANVACAQLFPSSYSSDN 155
 DB 1030 AVCOAPQYRLLEGGRNNEGRLVKKHGVGSCVDDDFNLSAQVACNSMGFEGPAKIEK 1089
 QY 156 LRVSSLEGQREFEVSIDHL---PDDKYALHHSYVVRGCSAGHYVTLQCTA----- 206
 DB 1090 NIFGNSNGP-----IWLDOVCFGENETSIDQCNHMMNGEHNCHTDEVALHCSAGPPRS 1144
 QY 207 -----CG-----HRGYS 214
 DB 1145 QRYSQIQIKGRSLGRFTTPKITYSQIGLMERSKAVHTPRRCGIFKDDLDTEVAHR---E 1201
 QY 215 SRIYVGNMSSLQMPWQASIQFOG-----YHLCGGSVITPLMITTAHCVYDLYLPK-SW 268

DR HSPB: P00750; 1.RTP.
DR FlyBase; FBgn0023479; Tegulia.
DR Interpro: IPR000194; ATPase_alpha_beta.
DR Interpro: IPR002557; Chitin_binding.
DR Interpro: IPR001314; Chitin_tyrpsin.
DR Interpro: IPR002172; LDL_recept_A.
DR Interpro: IPR002965; P_rich_extensn.
DR Interpro: IPR001190; SRCR.
DR Interpro: IPR001254; Tyrpsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00057; Idl_recept_a; 2.
DR Pfam: PF00530; SRCR; 2.
DR Pfam: PF00089; tyrsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR00258; SPBRCTRCPTR.
DR SMART: SM00494; CnLBD; 2.
DR SMART: SM00192; Ldla; 2.
DR SMART: SM00202; SR; 2.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS01209; LDLA_1; 1.
DR PROSITE; PSS0060; LDLA_2; 2.
DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
DR PROSITE; PSS0287; SRCR_2; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

Best Local Similarity 27.58; Pred. No. 2.5e-41;

Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

```
OY 66 GLGIH-----FDCKSKY-RCRSSFKCIELARCDGVSDDCKGDEDEYRCV 108
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1905 GWCYHNGCVDEVAGYTCVPMKCPNNYWLCHTKCECIPAFVCDNTPDCADKSD--CA 1962
OY 109 RV-----GGQNA---VLOVFTASWKTMCSDPMKGHYANVACAOLGFPSPYSSDN 155
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1963 AVCOAPQYRLGGRNSNGRLEVKHHGVWGVCDDDFNLKSAQVACNSMGCFGPATIEK 2022
OY 156 LRVSSLEGQFREEFVSIIDHL---PDDKYTALHHSYVREGCASGHVYTLQCTA----- 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2023 NIFGNSNGP-----IWLDOYWCFGNETSIDQCNHNMNGEHNCHTEDYALHCSAGPPPRS 2077
OY 207 -----CG-----HRRGYS 214
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2078 QRSQTOIKGGRSLGRRTTPTKYSQIGLMERSKAAVHTPRRCGTFKDDLTDEYAH--E 2134
OY 215 SRIVGMMSLISQMPQASLOFQG----YHLCGSVITPLWITTAACVYDLYLPK-SW 268
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2135 ERVVRGNVAORGRHPQATITRTGRGCISSHMCQAVVYSKRHLTLTAHCLYG--SPKAY 2192
OY 269 TIQVG--LVSLDNPAPSHLVEKIVYHSKY-KPKRLGNDIALMKLAGPLTFNEMIOPVCL 325
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2193 FVRGDDHYANIAESSEVDSEFTENNYLHENFRKGTMMNNDIALVVLKTPLEFSDYVOPICL 2252
OY 326 PNSEENPDDGKVCWTSGGATEDEGADASPYLNHAAVPLISNKCINHRDVGIIISPML 385
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2253 PDKNAELVEDRKCTISGMSIKSGVSTPAQVYLSAELPILADHYCKOSNYGSAWSEGMF 2312
OY 386 CAGYLITGGVDSGCGDSSGPLYVCOERRLMKLVGATSFSGICAEVKNPGVYTRVTSFLDMIH 445
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2313 CAGSMDESVDACEGDSGPLYVCSDDDEFTLYGLISMCGHCGFKNRPgyVYVNVNHYIDMiy 2372
OY 446 EOMERDL 452
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2373 EKINESL 2379
```

Search completed: April 11, 2002, 08:54:33
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 19.48 Seconds
(without alignments)
1775.321 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447
Sequence: 1 MGENDEPPAVEAPFSPRSIFG.....TRVTSFLDWIHEQMERDLKT 454

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

- Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	664.5	27.2	417	1 S00845	hepsin (EC 3.4.21.
2	660	27.0	1035	1 A43090	enteropeptidase (E
3	657	26.8	1019	1 A56318	enteropeptidase (E
4	654	26.7	1034	1 A53663	enteropeptidase (E
5	645	26.4	416	1 A53777	hepsin (EC 3.4.21.
6	570	23.3	638	1 KORTPL	plasma kallikrein
7	561	22.9	638	1 KOMSPL	plasma kallikrein
8	540	22.1	638	1 KOHUP	plasma kallikrein
9	526	21.5	790	1 PLPG	plasma kallikrein
10	523.5	21.4	761	1 JC5759	plasma (EC 3.4.21
11	520.5	21.3	455	2 A61545	brain-specific ser
12	507.5	20.7	812	1 PLMS	plasma (EC 3.4.21
13	507	20.7	437	2 S18407	plasma (EC 3.4.21
14	506	20.7	1113	2 JE0315	low-density lipopr
15	504.5	20.6	810	1 PLHU	plasma (EC 3.4.21
16	503.5	20.6	273	2 A47246	tryptase (EC 3.4.2
17	503.5	20.6	436	2 JX0172	tryptase (EC 3.4.21
18	503	20.6	460	2 B61545	plasma (EC 3.4.21
19	501.5	20.5	810	2 I46260	plasma (EC 3.4.21
20	500.5	20.5	275	2 A32410	tryptase (EC 3.4.2
21	500.5	20.5	418	2 A37344	acrosin (EC 3.4.21
22	500.5	20.5	421	2 S11674	acrosin (EC 3.4.21
23	498.5	20.4	431	2 S47538	acrosin (EC 3.4.21
24	498.5	20.4	812	1 PLBO	plasma (EC 3.4.21
25	497.5	20.3	810	2 B30848	plasma (EC 3.4.21
26	495.5	20.2	270	2 S56160	mast cell tryptase
27	494.5	20.2	625	1 KFH01	coagulation factor
28	490.5	20.0	274	2 A38654	mast cell proteina
29	488.5	20.0	274	2 JC4171	tryptase (EC 3.4.2

30	486.5	19.9	343	1 A57014	prostatin (EC 3.4.
31	481.5	19.7	421	2 S29599	acrosin (EC 3.4.21
32	480.5	19.6	275	2 A35863	tryptase (EC 3.4.2
33	479.5	19.6	275	2 B35863	tryptase (EC 3.4.2
34	476	19.5	264	2 I38136	chymotrypsin-like
35	474.5	19.4	275	2 C35863	tryptase (EC 3.4.2
36	472.5	19.3	263	2 S47537	chymotrypsin (EC 3
37	471.5	19.3	263	1 KYRTB	chymotrypsin (EC 3
38	469.5	19.2	263	2 A31299	chymotrypsin (EC 3
39	468.5	19.1	263	2 A21195	chymotrypsin (EC 3
40	463.5	18.9	415	1 A34170	acrosin (EC 3.4.21
41	463.5	18.9	615	1 KFH012	coagulation factor
42	462.5	18.9	245	1 KYBOA	chymotrypsin (EC 3
43	462.5	18.9	420	2 A55283	acrosin (EC 3.4.21
44	462	18.9	274	2 A45754	tryptase (EC 3.4.2
45	462	18.9	1524	2 T30337	polypeptide - Afri

ALIGNMENTS

```
RESULT 1
S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: S00845
R:Leys, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A>Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane
A:Reference number: S00845; MUID:88209431
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LE>
A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C:Genetics:
A:Gene: GDB:HPN; TMPSR1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TM>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Dissulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 27.2%; Score 664.5; DB 1; Length 417;
Best Local Similarity 35.7%; Pred. No. 1,5e-45;
Matches 149; Conservative 62; Mismatches 143; Indels 63; Gaps 12;

QY 59 LILALALIGLGHFDCGSKRRCSSFKICIELARCQGVSDCKGDEDEYRCVRYGGONAVIQ 118
   ||| |||
Db 26 LILLALTAIG-----AASWAIVAVLLRSD-----QRLPYVQVSSADARIM 64
   ||| |||

QY 119 VE--FAASWKTKWCSDMKMGHVAIVACAOIGFPPSYSSONLRVSSLEGGFREFEVSIDHL 176
   ||| |||
Db 65 VEDKTEGTWRLLCSSRSNARVAGLSCDEMGFLRALTHSELDRVTRTGANGTSGFEVD-- 122
   ||| |||

QY 177 PDDKVTALHSHSVYRE-----GCASGHVTTLOCTACGHRGRGYSRIRVGNMISLQMPMQ 231
   ||| |||
Db 123 -----GRPLHPTORLLEIVSYVCCPRGRFLAICQCGRRKRLPYDVIAGRGROTSLSRMRMQ 177
   ||| |||

QY 232 ASLQFGYHLGCGSVITPLMITTAACHVYDLYPR-----SWTIQGVLSILDPADPSH 285
   ||| |||
Db 178 VSLRFGAHLGCGSLGSDWVLTAAHC-----FPERNRVLSRWVFAVAVA---QASPHG 229
   ||| |||

QY 286 L--VEKIVHSHKYPKR-----LGNDIALMKLAGPLTFNEMTOPVCLPNSSENFPPDG 336
   ||| |||
Db 230 LQLGQAVVYHCGYLPFDPDPNSSENSNDIALVHLSSPLPLEYTIQPVCLPAAGQALVGGK 289
   ||| |||
QY 337 VCMTSGMGATEDGAGDASPVLNHAAPLISNKCICNHRDVGGLTISPSMIGAGYLTVGVDS 396
   ||| |||
```


F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 26.8%; Score 657; DB 1; Length 1019;
Best Local Similarity 37.7%; Pred. No. 1.7e-44;
Matches 148; Conservative 67; Mismatches 148; Indels 30; Gaps 15;

67 LGIHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVR---VGGQNAVLOVPT 121
Db LGIPECKADHQCCKNG-ECVPLVNLCDHLEHEDSDADCFRFTNGTNNGLVRFRT 695
QY 122 AASWKTMSDDMKGHYANVACAOLGFPYSSDNLRVSSLEQGFREFFVSDHLLPDDKY 181
Db 696 OSIIMHTACAEWNTTQISNDVQCLLGLGTGNSSK--PIFSTDG--GPVKL-NTADG-- 747
QY 182 TALHSHVYV-REGCASHVYVLOCT--ACGHR---RGYSRIYGGNMSLLSOMPWASLOF 236
Db 748 ---HLLTPSQQLDLSRLLOCNHSKCKLAADITRKIVGSGNAKGAAMPVWGLY 804
QY 237 OGTHLCGGSVITPLMITTAHACVYDLYL-PSKWTIOVGL--VSLDNP-APSHLVEKIY 292
Db 805 GGLLCCGASLVSRDMLVSAHCVYGRNLEPSKWTALGLHMSNLTSPQTVRLIDELIY 864
QY 293 HSKYKPKRLGNDIALMKLAGPLTFNEMIOVCLPNSSENPDPGKVCMTSGWATEDGAD 352
Db 865 NPHYNNRRKNDIDAMHLEFKVNTDYIOPICLPEENQVFPGRNSISAGWG-TVYVQGT 923
QY 353 ASFVNLHAAPVPLISNKICNHR-DYGGIISPSMLCAGYLTGVDSCGDSGGLVCOERR 411
Db 924 TAILIOEADVPPLISNRCQOQMPREYN--ITENNICAGEEGIDSCGDSGGLMCOENN 981
QY 412 LMKLVGATSFSGICAEVKNKGVYTRVTSFLDNI 444
Db 982 RMLAGVTSFGYKCALPNNRPGVAYARVSTTEWI 1014

RESULT 4

A53663
N.Alterpeptidase (EC 3.4.21.9) precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
C:Accession: A53663
R.Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakel, N.; Tsukada, S.; Miki, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994
A>Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; MUID:94327548
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1034 <MAT>
A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from t
C:Comment: The mechanism of association with the membrane of the intestinal brush border
C:Keywords: glycoprotein; hydrolyase; serine proteinase; transmembrane protein; zymogen
F:12-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MC>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HC>
F:1199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:159-519/Domain: MAM homology <MAM>
F:541-646/Domain: C1r/C1s repeat homology <C1R>
F:658-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology <SRC>
F:800-1034/Product: enteropeptidase light chain #status predicted <LC>
F:800-1039/Domain: trypsin homology <TRY>
F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96

F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 26.7%; Score 654; DB 1; Length 1034;
Best Local Similarity 37.7%; Pred. No. 3e-44;
Matches 149; Conservative 65; Mismatches 147; Indels 34; Gaps 15;

67 LGIHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVR---VGGQNAVLOVPT 121
Db 652 LGIPECKADHQCCKNG-ECVLLVNLCDHLEHEDSDADCFRFTNGTNNGLVRFRT 710
QY 122 AASWKTMSDDMKGHYANVACAOLGFPYSSDNLRVSSLEQGFREFFVSDHLLPDDKY 181
Db 711 OSIIMHTACAEWNTTQISNDVQCLLGLGTGNSMPF-FSSGGG---PFVKLN----- 757
QY 182 TALHSHVYV-REGCASHVYVLOCT--ACGHR---RGYSRIYGGNMSLLSOMPWASLOF 234
Db 758 TAPNSGLILTASQCFEDSLILLOCNHSKCKKQVQAQEVSPKIVGNDREGAMPVWV 817
QY 235 OGTHLCGGSVITPLMITTAHACVYDLYL-PSKWTIOVGL--VSLDNP-APSHLVEKI 290
Db 818 YINGQLCCASLVSRDMLVSAHCVYGRNLEPSKWTALGLHMTSNLTSPQTVRLIDELI 877
QY 291 VHSKYKPKRLGNDIALMKLAGPLTFNEMIOVCLPNSSENPDPGKVCMTSGWATEDGA 350
Db 878 VINPHNNRRKNDIDAMHLEFKVNTDYIOPICLPEENQVFPGRNSISAGWG-IYO 936
QY 351 GDASPVNLHAAPVPLISNKICNHR-DYGGIISPSMLCAGYLTGVDSCGDSGGLVCOE 409
Db 937 GSPADILIOEADVPPLISNRCQOQMPREYN--ITENNICAGEEGIDSCGDSGGLMCL 994
QY 410 RMLKLVGATSFSGICAEVKNKGVYTRVTSFLDNI 444
Db 995 NNRMLAGVTSFGYKCALPNNRPGVAYARVSTTEWI 1029

RESULT 5

S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R.Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinas
A:Reference number: S33777; MUID:93305733
A:Accession: S33777
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:957928; PIDN:CAA50256.1; PID:957929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolyase; liver; serine proteinase; transmembrane protein
F:12-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,250-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 26.4%; Score 645; DB 1; Length 416;
Best Local Similarity 35.2%; Pred. No. 5.5e-44;
Matches 146; Conservative 61; Mismatches 156; Indels 52; Gaps 11;

60 IIALAIGLGHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVVGQNAVLOV 119
Db 17 VALATVGTLLPLTGIG---AASMAIVTILR-----SDQEPYQVQLSPGDSRLIV 65
QY 120 -PTAASWKTMSDDMKGHYANVACAOLGFPYSSDNLRVSSLEQGFREFFVSDHLLPDDKY 173
Db 66 DTEGTGTTSSRSNARVAGIGCEBMGFLRALHSELVPTAGANGTSGFCVDEGGCLP 125
QY 174 ---HLLPDDKVTALHSHVYVREGCASHVYVLOCTACGHRGYSRIYGGNMSLLSOMPW 230

Db 126 LAORLL--DVISVC-----DPRGRFLATCODGRRKRLPDRIVGAGDSSSLGRWP 175
 QY 231 QASLQFGYHGGSVYTPPLMTITAAHCYVDLILPK-----SWTIOGVSLINDNAPS 284
 Db 176 QVSLRDGTHLGGSLISGDMVLAHC-----PPENNRLSRKRVFAGAVARTSPHAYQ 230
 QY 285 HLEKTIYHSHKYPKR-----LGNDIALMKLAGPLTFENMIOPVCLPNSSENPDKVC 338
 Db 231 LGQAVAYTHGGYLPRNPPTIDENSNDIALVHSSLPLEIYQPCVLPAGQALVDSKVC 290
 QY 339 WISGWCATEDGADASVYLHAAVPLISNKCINHRDVGIIISPSMLCAGYLTGVDSCQ 398
 Db 291 TVTGMGNTQ-FYGGQAVVLQEARVPIISNEVCNSPDYGNQIKPKMFCAGYPEGIDACQ 349
 QY 399 GDSGGPLVCOER-----RLMKLVGATSFSGTCAGVKNKPGVYTRVTSFLDWMHEOM 449
 Db 350 GDSGGHVCEDRLISGTSRMRLCGIVSGTCALARKPGVYTKVYDFREMIFOAIK 404

RESULT 6 KORTPL

Plasma kallikrein (EC 3.4.21.34) precursor - rat
 N:Alternate names: Fletcher factor; kininogenin; serum kallikrein
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
 C:Accession: A39180; A33320; S06851; I53041; S06852
 R:Beaudien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mokay, M.; Chretien, M.; Seidah, N.G.
 Biochemistry 30, 1628-1635, 1991
 A>Title: Gene structure and chromosomal localization of plasma kallikrein.
 A:Reference number: A39180; MUID:91129236
 A:Accession: A39180
 A:Molecule type: DNA
 A:Residues: 1-638 <BEA>
 A:Cross-references: GB:J05315
 A>Note: the authors translated the codon GAG for residue 81 as Gln
 R:Seidah, N.G.; Ladeheim, R.; Mokay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur
 DNA 8, 563-574, 1989
 A>Title: The cDNA structure of rat plasma kallikrein.
 A:Reference number: A33320; MUID:90091743
 A:Accession: A33320
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-638 <SEI>
 A:Cross-references: GB:M30282; NID:q205010; PIDN:AAA41463.1; PID:q205011
 A>Note: part of this sequence, including the amino ends of both the heavy and light chain
 R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
 Biochim. Biophys. Acta 999, 103-110, 1989
 A>Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
 A:Reference number: S06851; MUID:90089457
 A:Accession: S06851
 A:Molecule type: protein
 A:Residues: 20-45;391-413 <PAQ>
 R:Seidah, N.G.; Ladeheim, R.; Mokay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur
 DNA Cell Biol. 8, 563-574, 1989
 A>Title: The cDNA structure of rat plasma kallikrein.
 A:Reference number: I53041
 A:Accession: I53041
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <RES>
 A:Cross-references: GB:M58590; NID:q206721; PIDN:AAA42069.1; PID:q206722
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
 are linked by one or more disulfide bonds.
 C:Genetics:
 A:Gene: PK
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SI>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <MATI>
 F:20-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
 F:291-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
 F:127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predict
 F:396/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 570; DB 1; Length 638;
 Best Local Similarity 40.8%; Pred. No. 9e-38;
 Matches 116; Conservative 52; Mismatches 88; Indels 28; Gaps 9;

QY 189 YVREGCASHVTL-----OCACGHRGYSRIYGMNLSLQPMQASLOFO---G 238
 Db 362 YEAG--SSGTSRLCKVSSDCTT-----KINRIYGVGNSSLGEPMQVSLQVLSQ 415
 QY 239 YHLOGSVITPLMTITAAHCYVDLILPKSWTIOGVLSL--LDNPAESHLEKTIYHSHK 296
 Db 416 NHMGGSILGRWILFRAHCFDGIYPDPVWRIYGIILNLEITNKPFSSIKELIHQX 475
 QY 297 KPRRLNDIALMKLAGPLTFENMIOPVCLPNSSENPDKVCWTSNGATDAGADASP 356
 Db 476 KMSGSDYDLALIKQPLNTEFOKPLPSKADNTIYTNCWVGTGYTRK-RGETONT 534
 QY 357 LNHAAVPLISNKC--NHRDVGGIISPSMLCAGYLTGVDSCGDSGCPVCOERLWK 414
 Db 535 LQRTITPLVNECQKKRYR---VITKMICAGYKKGSGIDACKGDSGCPVLCYKHSRWQ 591
 QY 415 LVGATSFSGTCAGVKNKPGVYTRVTSFLDWMHEOM---ERDLKT 454
 Db 592 LVGTSWEGCARKQEPGVYTKVAEYIDWILEYIOSSKERALET 635

RESULT 7

KOMSPL
 plasma kallikrein (EC 3.4.21.34) precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
 C:Accession: A36557
 R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Roche
 DNA Cell Biol. 9, 737-748, 1990
 A>Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compar
 A:Reference number: A36557; MUID:91090844
 A:Accession: A36557
 A:Molecule type: mRNA
 A:Residues: 1-638 <SEI>
 A:Cross-references: GB:M58588; NID:q200358; PIDN:AAA63393.1; PID:q200359
 A>Note: part of this sequence, including the amino ends of both the heavy and light c
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
 are linked by one or more disulfide bonds.
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; in
 F:1-19/Domain: signal sequence #status predicted <SI>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
 F:20-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
 F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 22.9%; Score 561; DB 1; Length 638;
 Best Local Similarity 34.8%; Pred. No. 4.7e-37;
 Matches 126; Conservative 61; Mismatches 109; Indels 66; Gaps 12;
 QY 132 DMKGHYANVACAOLGFPYSVSSDNLKRVSLGQFREFVSIIDHLPDDKVTALHSHVYVA 191


```

Db 301 DFGGEELNV-----TFVGADVCOECTKTRICQFF-IYSLLPD-----CKE 342
QY 192 EGC-----ASGHVYTL-----QCTACGRGYSRTYGMSSLLSQW 223
Db 343 EGGKCSLRISTDGSPPRIYGMOGSSGSLRLCKLVSDPCTT-----KINARIYGGTNA 397
QY 224 LLSQMPQASLQFQ---GYHLICGGSVITPMTITTAHCYVDLYLPKSWTIQVLSLV--L 278
Db 398 SLGEMPMQVSLQYKLVSGHLCGGSIIIGROWVLTAAHCFEGIRPYDVKRIYGLISLSEI 457
QY 279 DNPAPSHLVKIVYHSKYKPKRLGNDIALMKLAGPLTFENMIQVCLPNSSENFPGKVC 338
Db 458 TKETPSRIKELIHOEYKSEGNVDIALIKLQPLNTEFQKPICLPSKADNTIYITNC 517
QY 339 WTSGMKGTEGADGASPVLNHAAPLISNKC--NHRDVGGIISPSMLCAGLTGVDSD 396
Db 518 WVTGMGTYKE--QGETONILQKATIPLVPMNECQKKYRDY---VINKQMLCAGKEGDTA 573
QY 397 CGDGSGLPVCQERRLMKLVGATSFSGICAEVKKPGVYTRVTSFLDMIHQME---RDL 452
Db 574 CKDSSGSLPVCXKHSGRMOLVGTISWEGCGRKQDPGVYTVSEYMDILEKTOSSDVRL 633
QY 453 KT 454
Db 634 ET 635

```

RESULT 8

KOHUP
 plasma kallikrein (EC 3.4.21.34) precursor - human
 N:Alternate names: kininogenin; plasma prekallikrein
 C:Species: Homo sapiens (man)
 C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
 C:Accession: A00921, A37939
 R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
 Biochemistry 25, 2410-2417, 1986
 A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
 A:Reference number: A00921; MUID:86243359
 A:Accession: A00921
 A:Molecule type: mRNA
 A:Residues: 1-638 <CHU>
 A:Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:9190263
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2050-2056, 1991
 A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
 A:Reference number: A37939; MUID:91152016
 A:Accession: A37939
 A:Molecule type: protein
 A:Residues: 20-37:40-46, 'X', 48, 'H', 50, 'X', 52-70, 'H', 75-76, 'X', 78-80, 103-113, 131-140, 141-
 280-283, 'X', 285, 287-291, 'X', 293-295, 314-317, 'X', 319-320, 321-324, 'X', 329-333, 334-339, 'X',
 523-538-551, 562, 'X', 564-567, 573, 'X', 575-576, 578-583, 'X', 585, 592-604 <MCN>
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
 are linked by one or more disulfide bonds.
 C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r
 ingen and may also play a role in the renin-angiotensin system by converting prorenin i
 C:Genetics:
 A:Gene: GDB:KLK3
 A:Cross-references: GDB:127575; OMIM:229000
 A:Map position: 4q35-4q35
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-638/Product: plasma kallikrein #status predicted <MAT>
 F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
 F:20-109/Domain: apple repeat <API>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104, 47-77, 51-57, 111-194, 137-166, 141-147, 201-284, 227-256, 231-237, 292-375, 322-328, 383

F:127,308,396,453,494/Binding site: carboxydrate (Asn) (covalent) #status experimenta
 F:318-347,340-345/Disulfide bonds: #status predicted
 F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 22.1%; Score: 540; DB 1; Length 638;
 Best Local Similarity 35.7%; Pred. No. 2,3e-35;
 Matches 124; Conservative 57; Mismatches 116; Indels 50; Gaps 12;

```

QY 132 DMKGHYANVACAOGLGPPVYSSDNLRYSSLEGGFREFEVSIDHLDPD-----179
Db 301 DFGGEELNV-----TFVGADVCOECTKTRICQFFIYSLLPD-----CKE 342
QY 180 ----KYTALHSHYVYREGGASGHVYTL-----QCTACGRGYSRTYGMSSLLSQW 228
Db 352 SMDGSPRI--AYGTQG--SSGYSRLRCNTGDSVCTT-----KSTRIRYGGTNSWGEW 402
QY 229 PMQASLQFQ---GYHLICGGSVITPMTITTAHCYVDLYLPKSWTIQVLSLV--NPAP 283
Db 403 PMQVSLQVYKLVTAQRHLCGGSIIIGROWVLTAAHCFEGIRPYDVKRIYGLISLSEI 457
QY 284 SHLVKIVYHSKYKPKRLGNDIALMKLAGPLTFENMIQVCLPNSSENFPGKVCWTSGW 343
Db 463 FSOIKETIIHQNKYSEGNVDIALIKLQPLNTEFQKPICLPSKADNTIYITNCWTCW 522
QY 344 GATEDGAGDASPVLNHAAPLISNKC--NHRDVGGIISPSMLCAGLTGVDSD 396
Db 523 GFSEKE--KGETONILQKATIPLVPMNECQKKYRDY---VINKQMLCAGKEGDTA 573
QY 403 GPLVCOERRLMKLVGATSFSGICAEVKKPGVYTRVTSFLDMIHQME---RDL 452
Db 580 GPLVCKHNGMWRMLVGTISWEGCGARQDPGVYTVSEYMDILEKTOSSDVRL 633

```

RESULT 9

PLPG
 plasmin (EC 3.4.21.7) precursor - pig (fragment)
 N:Alternate names: plasminogen
 N:Contains: miniplasminogen
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997
 C:Accession: S03733; S03737; A25834
 R:Schaller, J.; Marti, T.; Roesseler, S.J.; Kaemper, U.; Rickli, E.E.
 Fibrinolysis 1, 91-102, 1987
 A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
 A:Reference number: S03733
 A:Accession: S03733
 A:Molecule type: protein
 A:Residues: 1-560 <SCH>
 R:Brundisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lerger, W.; Manneberg,
 Eur. J. Biochem. 114, 465-470, 1981
 A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma
 A:Reference number: S03735; MUID:81212097
 A:Accession: S03737
 A:Molecule type: protein
 A:Residues: 1-57 <BRU>
 R:Marti, T.; Schaller, J.; Rickli, E.E.
 Eur. J. Biochem. 149, 279-285, 1985
 A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen
 A:Reference number: A25834; MUID:85203907
 A:Accession: A25834
 A:Molecule type: protein
 A:Residues: 450-790 <MAR>
 C:Function:
 A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
 us the walls of the graafian follicle; also activates the urokinase-type plasminogen
 A:Pathway: fibrinolysis
 C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homol
 C:Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine pr
 F:1-790/Product: plasminogen #status predicted <PEO>
 F:1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>
 F:1-77/Domain: activation peptide #status predicted <APM>

```

QY 70 HFD-----CSGKYRCSRFFCIELLARCDGVSDDCKDGED-----103
    |||      |||  ||  |  |  ||  ||
Db 326 HFDGSNRPIMLDVSCSGK---EVSF--IQCSRQGWGRHDCSHREDVGLTGYPPDSGHL 380

```

Db 204 QCESPFDCGKRVKPKKCSGIVGCGVAIAHSMFQISLRFRFGRIHFCGGTLLISPEVL 263

Page 6

```

F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match          20.7%; Score 507.5; DB 1; Length 812;
Best Local Similarity 39.8%; Pred. No. 1,26-32;
Matches 106; Conservative 40; Mismatches 91; Indels 29; Gaps 7;

QY 194 CASGHVVTLOCTACGHRGSGYSSRIYGVGNMILLSQMPQASL--QEQVHLCGGSVITPLM 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 568 CGKPOVERPKK-----PGRVVGCCVADPHSHSPWIDISLRTFTGHHFGCGTLADPEW 618

QY 252 ITTAHCYVDYDLPKSMITQVG---LVSLLDNPAASHLVKRIYHSKRYKFRRLGNDIAL 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 VLTAAHCLKSSRPREFYKVLIIIGAHEEYIRGLDYQETIS--VAKLLEPN-----NRDIAL 670

QY 308 MKLACPLTFNEMIDQVCLPNSDENFPDCKVCWTSGMATED--GAGDASPLNHAAYPLI 365
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 LKLSPPATITTKVILPACLPSPMYADRTICYITYGWGTQTFGGR---LKEAQLPVI 726

QY 366 SNKICNHDDVYGIIISPSMLCAGYLTCGVDSQCGSGSPVLCQERLKLKIVGATSFSGIC 425
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 727 ENKCNRRREYLNIRKXKSELCAQGLAGVDSQGSQSPVLCFEDKTIILQCVTSMGLCC 786

QY 426 AEVNPQVYTRVTSFLDMIHQMERD 451
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 ARPKNPGYVYRVSRFVDWIEREMRNN 812

RESULT 13
S18407
acrosin (EC 3.4.21.10) precursor - rat
N:Contains: proacrosin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: S18407; S30037; A56620
R:Kleim, U.; Flake, A.; Engel, W.
Biochim. Biophys. Acta 1090, 270-272, 1991
A:Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic
A:Accession: S18407; M01D:92031708
A:Molecule type: mRNA
A:Residues: 1-437 <KLE>
A:Cross-references: EMBL:X59254
R:Kleim, U.; Flake, A.; Engel, W.
submitted to the EMBL Data Library, April 1991
A:Reference number: S30037
A:Accession: S30037
A:Molecule type: mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KLE2>
A:Cross-references: EMBL:X59254; NID:957282; PIDN:CA041947.1; PID:957283
R:Kleimling, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
DNA Seq. 2, 57-60, 1991
A:Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
A:Reference number: A56620; M01D:92199245
A:Accession: A56620
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KRE>
A:Note: sequence modified after extraction from NCBI backbone
A:Note: sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447,
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-43/Domain: acrosin light chain #status predicted <LCH>
F:44-286/Domain: trypsin homology <TRY>
F:44-437/Domain: acrosin heavy chain #status predicted <HCH>
F:22-211/Binding site: carboxylate (Asn) (covalent) #status predicted
F:25-155,29-163/Disulfide bonds: #status predicted
F:74-90/Disulfide bonds: #status predicted
F:89,143,241/Active site: His, Asp, Ser #status predicted
F:178-247/Disulfide bonds: #status predicted
F:210-226/Disulfide bonds: #status predicted
F:237-267/Disulfide bonds: #status predicted

```

```

F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match          20.7%; Score 507.5; DB 1; Length 812;
Best Local Similarity 39.8%; Pred. No. 1.2e-32;
Matches 106; Conservative 40; Mismatches 91; Indels 29; Gaps 7;

Qy 194 CASGHVVTLOCTACGHRGSGYSSRIYGVGNMILLSQMPQASL--QEQVHLCGGSVITPLM 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 568 CGKPOVERPKK-----PGRVVGCCVADPHSHSPWIDISLRTFTGHHFGCGTLADPEW 618

Qy 252 ITTAHCYVDYDLPKSMITQVG---LVSLLDNPAASHLVKRIYHSKRYKFRRLGNDIAL 307
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 619 VLTAAHCLKSKSRPEFYKVILLGAHEEYIRGLDYQETIS--VAKLLEPN-----NRDIAL 670

Qy 308 MKLACPLTFNEMIQDVCCLPNSDENFPDCKVCWTSGMATED--GAGDASPLNNAAYPLI 365
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 671 LKLSPPATITTKVILPACLPSPMYADRTICYITYGWGTQTFGGR---LKEAQLPVI 726

Qy 366 SNKICNHDDVYGGIISPMSMLCAGYLTCGVDSQCGSGSPVLCQERLMLKIVGATSFSGIC 425
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 727 ENKCNRRREYLNINRRKSKSELCAGLAGVDSQGSQSGPVLVCFERDKITLLQCVTSMGLCC 786

Qy 426 AEVNPQVYTRVTSFLDMIHQMERD 451
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 ARPKNPGYVYRVSREVDWIEMERNN 812

RESULT 13

S18407
acrosin (EC 3.4.21.10) precursor - rat
N:Contains: proacrosin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: S18407; S30037; A56620
R:KiEmm, U.; Flake, A.; Engel, W.
B:Biochim. Biophys. Acta 1090, 270-272, 1991
A:Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic
A:Accession: S18407; M01D:92031708
A:Molecule type: mRNA
A:Residues: 1-437 <KLE>
A:Cross-references: EMBL:X59254
R:KiEmm, U.; Flake, A.; Engel, W.
submitted to the EMBL Data Library, April 1991
A:Reference number: S30037
A:Accession: S30037
A:Molecule type: mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KLE2>
A:Cross-references: EMBL:X59254; NID:957282; PIDN:CA041947.1; PID:957283
R:KiEmm, U.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
DNA Seq. 2, 57-60, 1991
A:Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
A:Reference number: A56620; M01D:92199245
A:Accession: A56620
A:Status: preliminary
A:Molecule type: DNA, mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KRE>
A:Note: sequence modified after extraction from NCBI backbone
A:Note: sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447,
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-43/Domain: acrosin light chain #status predicted <LCH>
F:44-286/Domain: trypsin homology <TRY>
F:44-437/Domain: acrosin heavy chain #status predicted <HCH>
F:22-211/Binding site: carboxylate (Asn) (covalent) #status predicted
F:25-155,29-163/Disulfide bonds: #status predicted
F:74-90/Disulfide bonds: #status predicted
F:89-143,241/Active site: His, Asp, Ser #status predicted
F:178-247/Disulfide bonds: #status predicted
F:210-226/Disulfide bonds: #status predicted
F:237-267/Disulfide bonds: #status predicted

```

Query Match	20.7%;	Score 507;	DB 2;	Length 437;
Best Local Similarity	42.4%;	Pred. No. 6.3e-33;		
Matches 104;	Conservative 36;	Mismatches 89;	Indels 16;	Gaps 5

QY	216	RIVGNSMISLISGMPQASIQF-----OGYNLCGGSVITPMLWITPAHC-----YVDLY	263
Db	42	RIVGGOTSSRAMPMWVSLQIFTSNHSRXYHACGSGLLMSHWLTAHCEFDNKKKKYYDWR	101
QY	264	LP-KSWTIOVLVLSDINPAPSHLEVETIVYHSKYKPKRLGNDIALTKLGLPTFNMIOIP	322
Db	102	LVEFGHEIEYEGNKKPVKEPODERYVOKIVIHKKYNAVTGENDLALLKTYPPYCGFVGP	161
QY	323	VCLPNSSENFDP-GKYCWTSFGWATEDGADSPVLANHAAPVLSIKCNHNRDVGIIIS	381
Db	162	GCLPFGKSGPRPIPIHTCYVTGNGYIKDNAPRPSVULMEARVDDLDLDLCNSTOWYNGRYT	221
QY	382	PSMLCAGYLTCGVDSQCGDSGSPLYCQF--RLMLKLVGATSFQIGCAEYNNKPGVYTRYVTS	439
Db	222	STNYCAGPBEKIDPCGDSGSPILMCRIOTRROPFVIVIGITISWGVCARAKRPBGVYATATWD	281
QY	440	FLDWTI 444	
Db	282	YLDWTI 286	

RESULT 14
JE0315
low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
C:Accession: JE0315
R:Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A:Reference number: JE0315; MUID:96429596
A:Accession: JE0315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1113 <TOM>
A:Cross-references: DDBJ:AB013874; NID:g3869144; PIDN:BA34371.1; PID:g3869145
C:Superfamily: trypsin homology; LDL receptor ligand-binding repeat homology <LDL1>
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:869-1097/Domain: trypsin homology <TRY>

Query Match	20.7%;	Score 506;	DB 2;	Length 1113;
Best Local Similarity	31.7%;	Pred. NO. 2.2e-32;		
Matches 126;	Conservative 76;	Mismatches 160;	Indels 36;	Gaps 16

QY	72	DCS-----GKXKRSSEFCIELIARCDVSCCKGEDEYCVRY----GGQAVLQVTPAAS	124
Db	719	NCSCQDQNELEC-ANHECVPRDLMOGDWDCCSSDBMGCVTLSTKQNSSSLTYKRSK	777
QY	125	WKTSCDMDWKGHYANVACAOLGF--PSYVSSDNLRVSLSGEOREEFVSLDHLPLPDKYT	182
Db	778	EHNVCAADGWRETLSQLCKQMLGEPSEV-----TKLIGEGGQQLRLRYPMWENLNS	830
QY	183	ALHHSVYVREGCASHVYVTLQCTA--CGHRRG--YSRIVGGMNLSLSQMPWQASLQFO-	237
Db	831	TLOELVYRRHSCPSRSEISLSCQDGRPARAMNKRILIGRTSPRGKPMWQCSLQSEP	890
QY	238	GYNHCGSVITPLMIITAAHCYVDLYLKPSKWTIQUGVLSLLDMPA---PSHLYEKIVNHS	294
Db	891	SGHICGGVLLAKKKNVLLVAHCEGREDADAVKAVFG--TNLNDHPSGGMQRRYFKTLLHR	949

```
QY      295 KYPRKRGNIDIAIMKLAGLPTLFNEMIOPLCPSEENFPDGAKCMTSGGATEDACDAS 354
       + : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      950 RYSAVAVDIIVSVELSDDIINETSIVRPPCLPSPPEYLEPDPTYCYITGMG----HMGNM 1005
       + : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      355 PV-FNHAAVPLINSKIT-C-NHRDVYGGIIISPMSCAGYLFGVDSCOGSDGGLVCQE--RR 411
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1006 PFKIQEEVEVKIIPLEQCQSFEDM--KTIIIRMKACIESGYTVDSCMDGSDGFLVCEPFG 1063
       + : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      412 LMRVGATSFEGIC-AEVNPFGVYTVTSFLLDIHEQM 448
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1064 QWTLFGLTSMGVSFCPSVLDPGRYSNVSTFVGHIEKQI 1101
       + : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

PLNU	15	RESULT	15
Plasmin (EC 3.4.21.7) precursor [validated] - human			
N:Alternate names: plasminogen precursor [misnomer]			
N:Contains: angiotatin; microplasmin; plasminogen			
C:Species: Homo sapiens (man)			
C:Date: 24-Apr-1994 #sequence_revision:02-Dec-1994 #text_change:15-Sep-2000			
C:Accession: A35229; I52242; A26646; I62738; I84609; S03735; A00929; A04627; A04625;			
R:Petersen, T.E.; Mattzen, M.R.; Ichinose, A.; Davie, E.W.			
J. Biol. Chem. 265, 6104-6111, 1990			
A:Title: Characterization of the gene for	human plasminogen, a key proenzyme in the f		
A:Reference number: A35229; MUID:90202879			
A:Accession: A35229			
A:Molecule type: DNA			
A:Residues: 1-810 <PRT>			
A:Cross-references: GB:O05286; GB:M34276; NID:g190064; PIDN:AAA60113.1; PID:g387026			
A:Experimental source: Leukocyte; Lung fibroblast			
R:Malgarretti, N.; Birno, L.; Pontoglio, M.; Candiani, G.; Meroni, G.; Ottolenghi, S.;			
Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990			
A:Title: Definition of the transcription initiation site of human plasminogen gene in			
A:Reference number: I52242; MUID:91097523			
A:Accession: I52242			
A:Status: translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-16 <MAL1>			
A:Cross-references: GB:M62890; NID:g190092; PIDN:AAA36454.1; PID:g553613			
R:Forstren, M.; Raden, B.; Israelsson, M.; Larsson, K.; Heden, L.O.			
FEBS Lett. 213, 254-260, 1987			
A:Title: Molecular cloning and characterization of a full-length cDNA clone for human			
A:Reference number: A26646; MUID:87162490			
A:Accession: A26646			
A:Molecule type: mRNA			
A:Residues: 1-471,'D',473-810 <FOR>			
A:Cross-references: GB:X05199; NID:g35530; PIDN:CAA28831.1; PID:g35531			
A:Experimental source: Liver			
R:Malinowski, D.P.; Sadler, J.E.; Davie, E.W.			
Biochemistry 23, 4243-4250, 1984			
A:Title: Characterization of a complementary deoxyribonucleic acid coding for human a			
A:Reference number: I45961; MUID:85023311			
A:Accession: I62738			
A:Status: translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 292-471,'D',473-810 <MAL2>			
A:Cross-references: GB:X02922; NID:g190112; PIDN:AAA60124.1; PID:g387031			
A:Accession: I84609			
A:Status: translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 367-419 <MAL3>			
A:Cross-references: GB:X02921; NID:g190110; PIDN:AAA60123.1; PID:g190111			
R:Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg,			
Eur. J. Biochem. 114, 465-470, 1981			
A:Title: Completion of the primary structure of the N-terminal CNBr fragments of huma			
A:Reference number: S03735; MUID:81212097			
A:Accession: S03735			
A:Molecule type: Protein			
A:Residues: 20-71,'E',73-76 <BRU>			
R:Soltrup-Jensen, L.; Petersen, T.E.; Magnusson, S.			
submitted to the Atlas, July 1977			
A:Reference number: A00929			
A:Accession: A00929			

